

02/04/00  
jc682 U.S. PTO

02 - 07 00

# Gray Cary Ware ▲ Freidenrich LLP

4365 Executive Drive, Suite 1600, San Diego, CA 92121-2189  
Phone 858-677-1400 Fax 858-677-1465 www.graycary.com

A  
jc625 U.S. PTO  
09/498098  
02/04/00

PATENT

Attorney Docket No. AURO1330

Box PATENT APPLICATION  
Assistant Commissioner For Patents  
Washington, D.C. 20231

CERTIFICATE OF MAILING BY "EXPRESS MAIL"
"EXPRESS MAIL" MAILING LABEL NO <u>EL476990941US</u>
DATE OF DEPOSIT <u>February 4, 2000</u> , I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C.F.R. 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER FOR PATENTS, WASHINGTON, D.C. 20231
<u>MIKHAIL BAYLEY</u> (TYPED OR PRINTED NAME OF PERSON MAILING PAPER OR FEE)
<u>Mikhail Bayley</u> SIGNATURE OF PERSON MAILING PAPER OR FEE

Sir:

Transmitted herewith for filing is the NEW patent application of:

Inventors: **JEFFREY H. STACK, MICHAEL WHITNEY, ANDREW B. CUBITT, AND  
BRIAN A. POLLOK**

For: **METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF**

Enclosed are the following papers, including all those required for a filing date under  
37 CFR § 1.53(b):

**195** pages of the Specification which includes **103** pages of Sequence Listing, **11** pages of the  
claims; **1** page of the Abstract;

**11** sheets of Formal Drawings;

**5** sheets of Assignment including recordation form of the invention to **AURORA BIOSCIENCES  
CORPORATION**;

**3** pages of Signed Declaration;

**2** pages of Power of Attorney by Assignee;

Return Postcard; and

Total fee of **\$1,378.00** which includes filing fee of \$1,338.00 and Assignment recordation of  
\$40.00 to be charged to Deposit Account No. 07-1895.

FULL NAME OF FIRST INVENTOR	LAST NAME: <b>STACK</b>	FIRST NAME: <b>JEFFREY</b>	MIDDLE NAME: <b>H.</b>
CITIZENSHIP	STATE OR FOREIGN COUNTRY: United States		
POST OFFICE ADDRESS	POST OFFICE ADDRESS: 6619 Edmonton Avenue	CITY AND STATE: San Diego, CA	ZIP CODE: 92122
FULL NAME OF SECOND INVENTOR	LAST NAME: <b>WHITNEY</b>	FIRST NAME: <b>MICHAEL</b>	MIDDLE NAME: none
CITIZENSHIP	STATE OR FOREIGN COUNTRY: United States		
POST OFFICE ADDRESS	POST OFFICE ADDRESS: 5171 Frink Avenue	CITY AND STATE: San Diego, CA	ZIP CODE: 92117
FULL NAME OF THIRD INVENTOR	LAST NAME: <b>CUBITT</b>	FIRST NAME: <b>ANDREW</b>	MIDDLE NAME: <b>B.</b>
CITIZENSHIP	STATE OR FOREIGN COUNTRY: United States		
POST OFFICE ADDRESS	POST OFFICE ADDRESS: 11083 Caminito Encanto	CITY AND STATE: San Diego, CA	ZIP CODE: 92131
FULL NAME OF FOURTH INVENTOR	LAST NAME: <b>POLLOK</b>	FIRST NAME: <b>BRIAN</b>	MIDDLE NAME: <b>A.</b>
CITIZENSHIP	STATE OR FOREIGN COUNTRY: United States		
POST OFFICE ADDRESS	POST OFFICE ADDRESS: 18283 Sun Maiden Court	CITY AND STATE: San Diego, CA	ZIP CODE: 92127

1. Cancel in this application original claims 41-49, 51-54, 56-59, and 61-79.
2. A preliminary amendment is enclosed.

The filing fee has been calculated as shown below:

For	Number Filed		Number Extra		Rate			Fee	
					Small Entity	Other Entity		Small Entity	Other Entity
Total Claims	43	-20 =	23	X	\$09	\$18	=		414.00
Independent Claims	6	- 3 =	3	X	\$39	\$78	=		234.00
Multiple Dependent Claims Presented: <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No					\$130	\$260	=		.00
BASIC FEE					\$345	\$690	=		690.00
					TOTAL FEE				\$1338.00

3. The Assistant Commissioner is hereby authorized to charge TOTAL FEE of \$1,378.00 which includes \$1,338 filing fee and \$40.00 Recordation of one (1) Assignment, and any other associated fees that may be required with this paper, or credit any overpayment, to Deposit Account No. 07-1895. A copy of the Transmittal Sheet is enclosed.

Any additional filing fees required under 37 C.F.R. 1.16.

Any patent application processing fees under 37 C.F.R. 1.17.

Respectfully submitted,

Date: 2/4/00



Lisa A. Haile, Ph.D.  
Attorney for Applicant  
Registration No. 38,347  
Telephone: 858-677-1456  
Facsimile: 858-677-1465

GRAY CARY WARE & FREIDENRICH LLP  
4365 Executive Drive, Suite 1600  
San Diego, CA 92121-2189

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of: ) Group Art Unit: Unassigned  
Stack et al. ) Examiner: Unassigned  
Application No.: Unassigned ) CERTIFICATE OF MAILING BY "EXPRESS MAIL"  
Filed: Herewith ) "EXPRESS MAIL LABEL NUMBER. EL476990941US  
For: METHODS OF PROTEIN ) DATE OF DEPOSIT: February 4, 2000  
DESTABILIZATION AND USES ) I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED  
THEREOF ) WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST  
OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE  
INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT  
COMMISSIONER OF PATENTS, BOX PATENT APPLICATION,  
WASHINGTON, D.C. 20231  
)  
) MIKHAIL BAYLEY  
) (TYPED OR PRINTED NAME OF PERSON MAILING PAPER)  
) *Mikhail Bayley*  
) (SIGNATURE OF PERSON MAILING PAPER OR FEE)

Assistant Commissioner of Patents  
Washington, D.C. 20231

**PRELIMINARY AMENDMENT**

Sir:

Please amend the application as set forth below.

**IN THE CLAIMS**

Please cancel claims 41-49, 51-54, 56-59 and 61-79 without prejudice..

**REMARKS**

Applicants have canceled claims 41-49, 51-54, 56-59 and 61-79 without prejudice.  
Consideration and entry of the preliminary amendment are requested.



In re Application of:  
Stack et al.  
Application No.: Not Assigned  
Filed: Herewith  
Page 2

PATENT  
Attorney Docket No.: AURO1330

No fee is deemed necessary in connection with the filing of this paper. However, if any fee is required, the Commissioner is hereby authorized to charge the amount of this fee, or credit any overpayments, to Deposit Account No. 07-1895. A copy of the Transmittal Sheet is enclosed.

Respectfully submitted,

Date:

2/4/00

Lisa A. Haile

Lisa A. Haile, Ph.D.

Registration No. 38,347

Telephone: (858) 677-1456

Facsimile: (858) 677-1465

GRAY CARY WARE & FREIDENRICH LLP  
4365 Executive Drive, Suite 1600  
San Diego, California 92121-2189

**CERTIFICATE OF MAILING BY "EXPRESS MAIL"**

"EXPRESS MAIL LABEL NUMBER" **EL476990941US**

DATE OF DEPOSIT **February 4, 2000**

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER OF PATENTS, BOX PATENT APPLICATION, WASHINGTON, D.C. 20231

**MIKHAIL BAYLEY**

(TYPED OR PRINTED NAME OF PERSON MAILING PAPER)

(SIGNATURE OF PERSON MAILING PAPER OR FEE)

**APPLICATION**

for

**UNITED STATES LETTERS PATENT**

on

**METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF**

by

**JEFFREY H. STACK, MICHAEL WHITNEY,  
ANDREW B. CUBITT, and BRIAN A. POLLOK**

Sheets of Formal Drawings: **Eleven (11)**

Docket No.: **AURO1330**

Attorneys

Lisa A. Haile, Ph.D., J.D.  
Gray Cary Ware & Freidenrich  
4365 Executive Drive, Suite 1600  
San Diego, California 92121-2189

## FIELD OF THE INVENTION

The present invention is in the field of protein analysis and more particularly methods of destabilizing proteins and using the destabilized proteins for novel cell based assays.

5

## BACKGROUND OF THE INVENTION

While genomic programs provide ever more sophisticated information on the sequence and patterns of expression of mammalian genes, it is increasingly  
10 recognized that integrating this information into a functional model of how a cell works requires an understanding of how the protein products of expressed genes interact within the cell. Although we have made significant improvements in our ability to clone, sequence and analyze DNA sequences, our reciprocal abilities for studying RNA and protein molecules are significantly less facile or advanced.  
15 Furthermore, proteins themselves represent significantly more complex molecules in terms of composition, shape and activity compared to double stranded DNA. A central challenge facing workers in the field today is to understand out how a protein's activity and function within a cell are regulated and coordinated within the native physiological context.

20 Traditionally, genetic analysis has been used for determining the function of gene products and how they interact with other proteins within a common pathway. Unfortunately genetic analysis in vertebrate organisms is extremely time consuming and expensive. An alternative approach is to devise an assay system for a given protein and then screen for compounds that activate or inhibit its function. These  
25 compounds can be used to dissect the cellular pathways the protein functions in, as well as serving as potential compounds of therapeutic value.

Although there is tremendous interest in understanding the regulation and interactions of proteins within cells there are relatively few methods that are robust, simple to use, amenable to high throughput screening or can be used effectively  
30 within living cells. Furthermore in many cases where specific assays do exist these

are restricted in scope to individual enzymatic steps or to one or two defined pathways.

A need thus exists for sensitive methods of interfacing the functional modifications of proteins with optical signals that can be used detect and monitor these changes, for example for use in high throughput screening. In drug screening applications these assays can be applied to find useful compounds that are specific and selective for a particular protein or signal transduction or metabolic pathway.

Proteins may undergo a huge variety of post-translational modifications subsequent to their synthesis in the cell. In many cases these modifications can play critical roles in the functioning and stability of the modified proteins. For example, proteolysis, phosphorylation, covalent attachment of a lipid or lipid derivative, disulfide bond formation, glycosylation and oxidation all can have important functional effects. Many other examples also exist and may play important functional roles within a cell for defined proteins.

One approach to developing a generic assay capable of detecting these myriad post-translational modifications is to operatively couple these activities through a central pathway of protein modification that can be sensitively measured with a common reporter system. In the present invention, the inventors have recognized that by coupling post-translational activities to the stability of a high sensitivity reporter moiety it is possible to develop uniform cell based assays for a range of activities. Importantly these measurements are robust enough for high throughput screening applications, readily adaptable to a range of activities and provide cellular assays that provide information within a living cell.

In the present invention, post-translational activities can be measured by providing one or more constructs in which the activity to be measured influences the stability of a reporter moiety. In one embodiment, this may be achieved by providing a reporter moiety that is operatively coupled to a multimerized destabilization domain via a linking moiety. The linking moiety comprises a recognition motif for the target activity such that modification of the linker by the activity results in altered stability of the reporter moiety. If the reporter moiety is an enzymatic reporter gene the method provides a high sensitivity readout that is generally applicable to a range of activities

which are otherwise difficult to measure within living cells. The multimerized destabilization domain described herein provides a key advantage in the method because it enables the degree of destabilization to be predictably tuned to any activity level or intrinsic stability of the target protein or reporter moiety.

5           The regulation of protein stability is an area of particular interest because of its increased recognition as a key regulator of a protein's concentration and function in the cell. Although our knowledge of the factors that control protein stability have grown dramatically in recent years, it is clear that a variety of cellular pathways and environmental cues participate in and control a protein's fate. For example, mis-  
10   folding, proteolysis, oxidation and some conformational changes that expose significant surface hydrophobicity readily contribute to the recognition of a protein by the cellular machinery for protein degradation. The majority of cytoplasmic protein degradation involves the ubiquitination of the target protein followed by binding and degradation by the proteasome. (For review see Hershko and Ciechanover (1998)  
15   Annu. Rev. Biochem. 67 425-79)

          A key step in protein ubiquitination, and degradation, is recognition of the target protein by ubiquitin protein ligase or E3 enzyme. This class of enzymes is responsible for the covalent attachment of ubiquitin to the target protein via an amide isopeptide linkage to an  $\epsilon$ -amino group of one of the substrate protein's lysine  
20   residues. There are currently believed to be multiple families of E3 enzymes, additionally there is increasing evidence that some E3 proteins exist as multi-subunit protein complexes (Laney and Hochstrasser (1999) Cell 97 427-430). E3 proteins and their associated complexes are believed to be largely responsible for recognizing and ubiquitinating damaged proteins as well as specific destabilization domains present in  
25   target proteins. Once recognized, a protein target that has been modified by the addition of a single ubiquitin domain, becomes a substrate for further ubiquitination, either at different sites in the substrate protein, or through extension of the conjugated ubiquitin. This process can thus lead to a poly-ubiquitinated protein with numerous branched ubiquitin domains attached. Once poly-ubiquitinated, the protein is  
30   recognized with high affinity by the proteasome where it is degraded.

The addition of specific destabilization domains to a target protein has in some cases been demonstrated to destabilize that target protein. A key challenge in this area has been to provide a predictable way of creating graded levels of destabilization for a given protein that that can be utilized in manipulating the steady state levels or dynamic temporal regulation of that protein. The present inventors have discovered for the first time that by providing stable multimerized linear chains of individual destabilization domains, such as ubiquitin, it is possible to create a generic method of protein destabilization that is widely applicable to virtually any protein. Importantly, this approach has the advantage that the degree of destabilization can be accurately controlled by varying the number of destabilization domains added to the target protein. As a result, the actual cellular concentration and half-life of an exogenously expressed protein in a cell or living organism can be accurately and reproducibly controlled. By coupling 1, 2, 3, 4 or more copies of ubiquitin to the reporter gene  $\beta$ -lactamase it has been possible to regulate the protein concentration of this protein in the cell over a 10-fold range compared to the native protein. The present inventors have applied this discovery to create an assay technology that is broadly capable of measuring a wide range of post-translational activities.

## SUMMARY OF THE INVENTION

This invention provides a fluorescent, bioluminescent or enzymatic substrate useful as an optical probe or sensor of post-translational modifications, such as proteolysis. In one embodiment, the invention comprises a reporter moiety that is functionally coupled to one or more destabilizing domains via a linker. The linker typically contains a recognition motif for an activity. Modification of the linker by the activity results in uncoupling of the reporter moiety from the destabilizing domain(s) with a corresponding change in the stability of the reporter moiety. The level of activity within a sample is sensed by a measurable change in the level of the reporter moiety, for example by detecting at least one optical property of the reporter moiety, or by detecting at least one optical property of detectable product of the reporter moiety. **FIG. 1.**

In one embodiment the reporter moiety is an enzymatic reporter such as alkaline phosphatase,  $\beta$ -galactosidase, chloramphenicol acetyltransferase,  $\beta$ -glucuronidase, peroxidase,  $\beta$ -lactamase, bioluminescent proteins, luciferases and catalytic antibodies. In another embodiment the reporter moiety is a naturally  
5 fluorescent protein, epitope or structural protein.

In one aspect the linker moiety is an amino acid sequence that covalently couples the reporter moiety to the multimerized destabilization domain. In another aspect, the linker moiety comprises two separate amino acid sequences, one of which is covalently coupled to the reporter moiety, and one of which is coupled to the  
10 multimerized destabilization domain. Coupling of the reporter moiety to the destabilization domains occurs through the non-covalent interaction or binding of the two amino acid sequences of the linker together. In either case, modification of the linker by the activity results in a modulation of the coupling of the reporter moiety to the multimerized destabilization domains. In one aspect of this method the activity is  
15 selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.

In one aspect the multimerized destabilization domain comprises two, three, four, or more copies of the destabilization domain covalently coupled together in a linear chain. In one embodiment, the destabilization domains comprise ubiquitin, or a  
20 homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In another aspect the invention comprises a method of regulating the  
25 concentration of one or more target proteins in a cell. The method involves the creation of a fusion protein containing the protein of interest coupled to one or more destabilization domains. In different embodiments the protein of interest may be coupled to a multimerized destabilization domain comprising two or more copies of  
30 the destabilization domain. In one embodiment, the destabilization domains comprise

ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In one aspect of this method, the fusion protein may additionally comprise a linker that couples the protein of interest to one or more destabilization domains. The linker typically comprises a protease cleavage site for a protease. Cleavage of the linker by the protease modulates the coupling of the multimerized destabilization domain to the protein of interest, thereby providing a method of rapidly modulating the stability of one or more proteins of interest in the cell simultaneously. The protease may be introduced into the cell, or its activity regulated by the presence of a membrane permeant small molecule inhibitor. In one embodiment of this method, the protease does not naturally occur in the target cell.

In another aspect the invention includes a recombinant DNA molecule, comprising a nucleic acid sequence encoding for one or more destabilization domains, a target protein, and a linker moiety that operatively couples the destabilization domain(s) to the target protein. In different embodiments the protein of interest may be coupled to one, two, three, four or more copies of the destabilization domain. In one embodiment, the destabilization domains comprise ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In another embodiment the invention includes a recombinant protein molecule, comprising an amino acid sequence encoding for one or more destabilization domains, a target protein, and a linker moiety that operatively couples the multimerized destabilization domain to the target protein.



In another aspect the invention includes a cell or transgenic organism comprising a nucleic acid sequence encoding for a one or more destabilization domains, a target protein, and a linker moiety that operatively couples the destabilization domain(s) to the target protein. In different embodiments the protein of interest may be coupled to one, two, three, four or more copies of the destabilization domain. In one embodiment, the destabilization domains comprise ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In another embodiment the invention includes a method for identifying a modulator of an activity, comprising the use of the inventions cells or transgenic organisms. The method includes contacting the cells with a test chemical and detecting the activity of the reporter moiety. Additional claims involve the steps of contacting the cell with an activator of the activity prior to the addition said test chemical, and of in parallel determining the cell viability of the cell in the presence of the test chemical.

In another embodiment the invention is directed to the test chemical and a pharmaceutical composition comprising a test chemical identified by the methods of the present invention.

The accompanying drawings, which are incorporated in and form part of the specification, merely illustrate embodiments of the present invention. Together with the remainder of the specification, they are meant to serve to explain certain principles of the invention to those of skill in the art.

## BRIEF DESCRIPTION OF THE FIGURES

**FIG. 1** General schematic overview of parent construct pcDNA3-UbiquitinG76V-Bla. Shown are important coding regions including the ubiquitin- $\beta$ -lactamase fusion

coding region, various promoters and important restriction sites used in the cloning of derivative constructs.

**FIG. 2** TNT *in vitro* synthesis and degradation experiments with Met, 1, 2, 3 or 4  
5 copies of ubiquitinG76V fused to  $\beta$ -lactamase. The kinetics of turnover *in vitro* in  
(A) were determined by chase reactions at 37°C and products analyzed by SDS-  
PAGE. The effect of the proteasome inhibitor MG132 at 50  $\mu$ M in the TNT synthesis  
reaction is shown in (B).

10 **FIG. 3** Turnover *in vitro* of labeled fusion proteins of uncleavable ubiquitinG76V  
fused to GFP. TNT synthesis reactions were incubated in chase lysate at 37°C and  
products analyzed by SDS-PAGE.

**FIG. 4** Turnover reactions *in vitro* of labeled uncleavable ubiquitin caspase-3 fusions.  
15 TNT reactions were incubated in chase lysate at 37°C and products analyzed by SDS-  
PAGE.

**FIG. 5** FACS™ analysis of uncleavable ubiquitin  $\beta$ -lactamase fusions. Jurkat cells  
expressing ubiquitinG76V-Bla fusion proteins were analyzed for  $\beta$ -lactamase  
20 expression by flow cytometry. The R5+R6+R7 region was designated as Bla<sup>+</sup> and the  
percentage of cells in that region is shown in the bar graph.

**FIG. 6** Kinetics of degradation *in vivo* of ubiquitinG76V- $\beta$ -lactamase fusion proteins.  
Jurkat cells expressing the various ubiquitinG76V-Bla fusions were treated with  
25 cycloheximide to initiate a chase and aliquots of cells were removed at the indicated  
times. The cells were lysed and the  $\beta$ -lactamase activity in the lysates was  
determined by an *in vitro* reaction using the fluorescent substrate CCF2. The  $\beta$ -  
lactamase activity was measured by cleavage of CCF2 and represented as emission at  
460nm.

30

**FIG. 7** Caspase cleavage of 2XUb-DEVD-Bla results in the stabilization of  $\beta$ -lactamase. TNT synthesis reactions were performed to generate labeled fusion proteins of the caspase substrate 2XUb-DEVD-Bla and control 2XUb-DEVA-Bla. In (A), the labeled proteins were incubated with purified caspase-3 to verify that 2XUb-DEVD-Bla can be cleaved by caspase-3 and 2XUb-DEVA-Bla cannot. In (B), the products of the caspase-3 cleavage reactions were incubated with chase extract and samples analyzed by SDS-PAGE.

**FIG. 8** Dose-response curves for an inducer and an inhibitor of caspase activation with Jurkat cells expressing 2XUb-DEVD-Bla. Varying concentrations of antiFas IgM were incubated with 2XUb-DEVD-Bla-expressing Jurkat cells for 6 hours at 37°C and caspase activity was measured following a cycloheximide chase to clear uncleaved reporters. The cells were loaded with CCF2-AM and  $\beta$ -lactamase activity measured and expressed as a 460/530nm ratio. Jurkat cells expressing 2XUb-DEVD-Bla were treated with varying concentrations of the caspase inhibitor ZVAD-fmk and then treated with 75 ng/ml antiFas IgM. The cells were incubated for 6 hours at 37°C, cycloheximide for 1 hour at 37°C and  $\beta$ -lactamase activity measured using CCF2-AM as described above.

**FIG. 9** *In vitro* cis-cleavage activity of UbG76V-HRV 2A-Bla fusions. Labeled UbiquitinG76V-HRV 2A protease  $\beta$ -lactamase fusions were produced in TNT reactions and then analyzed by SDS-PAGE. (A) shows that the cis-cleavage of HRV-Bla fusions is blocked by mutation of putative catalytic residues (C106 and D35). (B) The TNT reactions were incubated in chase extract to show the selective stabilization of the cleavage product.

**FIG. 10** Rapid degradation of 2XUb-Bla *in vitro* requires polyubiquitination and proteasome activity. TNT synthesis reactions were incubated in chase extract containing the indicated inhibitors for 20 minutes at 37°C. MG132 and ALLN were present at 50  $\mu$ M, lactacystin at 10 mM and MeUb at 200  $\mu$ g/ml.

**FIG. 11** Dose-response curves for proteasome inhibitors on Jurkat cells expressing 2XUb-Bla reporter. Cells were treated with varying concentrations of MG132 or ALLN for 30 minutes and then cycloheximide was added and the cells incubated at 37°C for one hour. The cells were loaded with CCF2-AM to measure  $\beta$ -lactamase activity as described above.

## DETAILED DESCRIPTION

### Definitions

The techniques and procedures are generally performed according to conventional methods in the art and various general references. (Lakowicz, J.R. Topics in Fluorescence Spectroscopy, (3 volumes) New York: Plenum Press (1991), and Lakowicz, J. R. (1996) Scanning Microsc Suppl. 10 213-24, for fluorescence techniques; Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for molecular biology methods; Cells: A Laboratory Manual, 1<sup>st</sup> edition (1998) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for cell biology methods; Optics Guide 5 Melles Griot® Irvine CA, and Optical Waveguide Theory, Snyder & Love published by Chapman & Hall for general optical methods, which are incorporated herein by reference which are provided throughout this document).

“Activity” refers to the enzymatic or non-enzymatic activity capable of modifying an amino acid residue or peptide bond (preferably enzymatic). Such covalent modifications include proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation and ADP-ribsoylation. The term includes non-covalent modifications including protein-protein interactions, and the binding of allosteric, or other modulators or second messengers such as calcium, or cAMP or inositol phosphates to a polypeptide.

Amino acid "substitutions" are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has similar structural and/or chemical properties. Examples of conservative replacements

are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine.

Amino acid "insertions" or "deletions" are changes to or within an amino acid sequence. They typically fall in the range of about 1 to 5 amino acids. The variation  
5 allowed in a particular amino acid sequence may be experimentally determined by producing the peptide synthetically or by systematically making insertions, deletions, or substitutions of nucleotides in the gene sequence using recombinant DNA techniques.

"Animal" as used herein may be defined to include human, domestic (cats,  
10 dogs, etc), agricultural (cows, horses, sheep, goats, chicken, fish, etc) or test species (frogs, mice, rats, rabbits, simians, etc).

"Chimeric" molecules are polynucleotides or polypeptides which are created by combining one or more of nucleotide sequences of this invention (or their parts) with additional nucleic acid sequence(s). Such combined sequences may be  
15 introduced into an appropriate vector and expressed to give rise to a chimeric polypeptide which may be expected to be different from the native molecule in one or more of the following characteristics: cellular location, distribution, ligand- binding affinities, interchain affinities, degradation/turnover rate, signaling, etc.

The terms "cleavage site" or "protease site" refers to the bond cleaved by the  
20 protease (e.g. a scissile bond) and typically the surrounding three to four amino acids of either side of the bond.

"Control elements" or "regulatory sequences" are those non-translated regions of the gene or DNA such as enhancers, promoters, introns and 3' untranslated regions which interact with cellular proteins to carry out replication, transcription, and  
25 translation. They may occur as boundary sequences or even split the gene. They function at the molecular level and along with regulatory genes are very important in development, growth, differentiation and aging processes.

"Corresponds to" refers to a polynucleotide sequence that is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference  
30 polynucleotide sequence, or that a polypeptide sequence is identical to all or a portion of a reference polypeptide sequence. In contradistinction, the term "complementary

to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

5 "Derivative" refers to those polypeptides which have been chemically modified by such techniques as ubiquitination, labeling, pegylation (derivatization with polyethylene glycol), and chemical insertion or substitution of amino acids such as ornithine which do not normally occur in human proteins.

10 A "destabilization domain" refers to a protein, polypeptide or amino acid sequence that is capable of modulating the stability of a protein of interest when functionally coupled to the protein of interest. Examples of destabilizing domains include ubiquitin, PEST sequences, cyclin destruction boxes and hydrophobic stretches of amino acids. Preferred destabilization domains include ubiquitin and homologs thereof, particularly those comprising mutations that prevent, or  
15 significantly reduce, the cleavage of ubiquitin multimers by  $\alpha$ -NH-ubiquitin protein endoproteases. Examples of such mutations include the mutation of glycine 76 to another amino acid, particularly an amino acid selected from the group consisting of Ala, Leu, Ile, Phe, Tyr, Val, Met, Cys, His, Trp, Pro, Arg, Lys, Thr and Ser. Preferred is UbiquitinG76V.

20 A "detectable product" is a chemical moiety used for detecting a reporter moiety. They include, but are not limited to, radionuclides, enzymes, fluorescent, chemi-luminescent, or chromogenic agents. Detectable products associate with, establish the presence of, and may allow quantification of a particular nucleic sequence, amino acid sequence or reporter moiety. Preferred detectable products are  
25 retained within living cells and provide a fluorescence readout that is compatible with fluorescent activated cell sorting (FACS) analysis.

The term "engineered protease site" refers to a protease site that has been modified from the naturally existing sequence by at least one amino acid substitution.

30 The term "homolog" refers to two sequences or parts thereof, that are greater than, or equal to 85% identical when optimally aligned using the ALIGN program. Homology or sequence identity refers to the following. Two amino acid sequences

are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M.O., (1972) in Atlas of Protein Sequence and Structure 5, National Biomedical Research Foundation, 101-110, and Supplement 2 to this volume, pp. 1-10.

An "inhibitor" is a substance that retards or prevents a chemical or physiological reaction or response. Common inhibitors include but are not limited to antisense molecules, antibodies, antagonists and their derivatives.

"Isolated" refers to material removed from its original environment (e.g. the natural environment if it is naturally occurring), and thus is altered from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

The term "linker" or "linker moiety" refers to an amino acid, polypeptide or protein sequence that serves to operatively couple a reporter moiety to one or more destabilization domains. Linkers may comprise a single polypeptide chain that covalently couples the reporter moiety to the multimerized destabilization domain. Alternatively the linker may comprise two separate polypeptides. Typically the first polypeptide is covalently coupled to the reporter moiety, and the second polypeptide is covalently coupled to the multimerized destabilization domain. Generally the first and second polypeptides comprising the linker moiety in this embodiment are capable of interacting or associating such that the interaction or association operatively couples the reporter moiety to the multimerized destabilization domain. Preferably the

linker moiety is non-cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. Linkers may be of any size.

The term “modulates” refers to, either the partial or complete, enhancement or inhibition (e.g. attenuation of the rate or efficiency) of an activity or process.

5       The term "modulator" refers to a chemical compound (naturally occurring or non-naturally occurring), such as a biological macromolecule (e.g., nucleic acid, protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian, including human) cells or tissues. Modulators are evaluated for potential activity as inhibitors  
10 or activators (directly or indirectly) of a biological process or processes (e.g., agonist, partial antagonist, partial agonist, inverse agonist, antagonist, antineoplastic agents, cytotoxic agents, inhibitors of neoplastic transformation or cell proliferation, cell proliferation-promoting agents, and the like) by inclusion in screening assays described herein. The activity of a modulator may be known, unknown or partially  
15 known.

The term “multimerized destabilization domain” refers to at least two destabilization domains that are linearly coupled together. Preferred multimerized domains are non-cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases. The term does not include naturally occurring poly-ubiquitin chains in which the ubiquitin  
20 monomers are coupled together via isopeptide bonds attached to the  $\epsilon$ -amino group of lysine. The term also does not include naturally occurring multi-ubiquitin genes, are cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases to create ubiquitin monomers. The destabilization domains present in the multimerized destabilization domain are typically the same, but need not necessarily be identical.

25       “Naturally fluorescent protein” refers to proteins capable of forming a highly fluorescent, intrinsic chromophore either through the cyclization and oxidation of internal amino acids within the protein or via the enzymatic addition of a fluorescent co-factor. Typically such chromophores can be spectrally resolved from weakly fluorescent amino acids such as tryptophan and tyrosine.



"Naturally occurring" refers to a polypeptide produced by cells which have not been genetically engineered or which have been genetically engineered to produce the same sequence as that naturally produced. Specifically contemplated are various polypeptides that arise from post-transnational modifications. Such modifications of the polypeptide include but are not limited to acetylation, carboxylation, glycosylation, phosphorylation, lipidation, proteolytic cleavage and acylation.

An "oligonucleotide" or "oligomer" is a stretch of nucleotide residues which has a sufficient number of bases to be used in a polymerase chain reaction (PCR), a site directed mutagenesis reaction or a cassette to create a desired sequence element. These short sequences are based on (or designed from) genomic or cDNA sequences and are used to amplify, mutate or create particular sequence elements. Oligonucleotides or oligomers comprise portions of a DNA sequence having at least about 10 nucleotides and as many as about 50 nucleotides, preferably about 15 to 30 nucleotides. They are chemically synthesized and may also be used as probes.

An "oligopeptide" is a short stretch of amino acid residues and may be expressed from an oligonucleotide. It may be functionally equivalent to and either the same length as or considerably shorter than a "fragment", "portion", or "segment" of a polypeptide. Such sequences comprise a stretch of amino acid residues of at least about 5 amino acids and often about 17 or more amino acids, typically at least about 9 to 13 amino acids, and of sufficient length to display biologic and/or immunogenic activity.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "operably coupled" refers to a juxtaposition wherein the components so described are either directly or indirectly coupled. Examples of directly coupled components include proteins that are translationally fused together. Examples of indirectly coupled components include proteins that can functionally associate either transiently, or persistently, through a binding interaction.

The term "polynucleotide" refers to a polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides. Modified forms and analogs of either type of nucleotide are also included, as are ribonucleotides or deoxynucleotides linked via novel bonds such as those described in U.S. Patent No. 5,532,130, European Patent Applications EP 0 839 830, EP 0 742 287, EP 0 285 057 and EP 0 694 559. The term includes single and double stranded forms of nucleotides, or a mixture of single and double stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine, as well as other chemical or enzymatic modifications.

The term "polypeptide" refers to amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modification include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pergylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to protein such as arginylation. (See Proteins- Structure and Molecular

Properties 2<sup>nd</sup> Ed., T.E. Creighton, W.H. Freeman and Company, New York (1993);  
Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic  
Pres, New York, pp. 1-12 (1983).

5 A "portion" or "fragment" of a polynucleotide or nucleic acid comprises all or  
any part of the nucleotide sequence having fewer nucleotides than about 6 kb,  
preferably fewer than about 1 kb which can be used as a probe. Such probes may be  
labeled with reporter molecules using nick translation, Klenow fill-in reaction, PCR  
or other methods well known in the art. After pretesting to optimize reaction  
conditions and to eliminate false positives, nucleic acid probes may be used in  
10 Southern, northern or in situ hybridizations to determine whether DNA or RNA  
encoding the protein is present in a biological sample, cell type, tissue, organ or  
organism.

"Probes" are nucleic acid sequences of variable length, preferably between at  
least about 10 and as many as about 6,000 nucleotides, depending on use. They are  
15 used in the detection of identical, similar, or complementary nucleic acid sequences.  
Longer length probes are usually obtained from a natural or recombinant source, are  
highly specific and much slower to hybridize than oligomers. They may be single- or  
double-stranded and carefully designed to have specificity in PCR, hybridization  
membrane-based, or ELISA-like technologies.

20 The term "recognition motif" refers to all or part of a polypeptide sequence  
recognized by a post-translational modification activity to enable a polypeptide to  
become modified by that post-translational modification activity. Typically, the  
affinity of a protein, e.g. enzyme, for the recognition motif is about 1 mM (apparent  $K_d$ ),  
preferably a greater affinity of about 10  $\mu$ M, more preferably, 1  $\mu$ M or most  
25 preferably has an apparent  $K_d$  of about 0.1  $\mu$ M. The term is not meant to be limited to  
optimal or preferred recognition motifs, but encompasses all sequences that can  
specifically confer substrate recognition to a peptide. In some embodiments the  
recognition motif is a phosphorylated recognition motif (e.g. includes a phosphate  
group), or comprises other post-translationally modified residues.

30 "Recombinant nucleotide variants" are polynucleotides that encode a protein.  
They may be synthesized by making use of the "redundancy" in the genetic code.

Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

5 "Recombinant polypeptide variant" refers to any polypeptide which differs from a naturally occurring polypeptide by amino acid insertions, deletions and/or substitutions, created using recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing characteristics of interest may be found by comparing the sequence of a polypeptide  
10 with that of related polypeptides and minimizing the number of amino acid sequence changes made in highly conserved regions.

A "reporter moiety" includes any protein that directly or indirectly produces a specific detectable product, or cellular phenotype, such as drug resistance that can be used to monitor transcription of a gene. Preferred reporter moieties include proteins  
15 with an enzymatic activity that provides enzymatic amplification of gene expression such as alkaline phosphatase,  $\beta$ -galactosidase, chloramphenicol acetyltransferase,  $\beta$ -glucuronidase, peroxidase,  $\beta$ -lactamase, bioluminescent proteins, luciferases and catalytic antibodies. Other reporter moieties include proteins such as naturally fluorescent proteins or homologs thereof, cell surface proteins or the native or  
20 modified forms of an endogenous protein to which a specific assay exists or can be developed in the future. Preferred reporter moieties for use in the present invention provide for a fluorescent readout that is compatible with fluorescent activated cell sorting (FACS) analysis.

A "signal or leader sequence" is a short amino acid sequence which is or can  
25 be used, when desired, to direct the polypeptide through a membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous sources by recombinant DNA techniques.

A "standard" is a quantitative or qualitative measurement for comparison. Preferably, it is based on a statistically appropriate number of samples and is created  
30 to use as a basis of comparison when performing diagnostic assays, running clinical

1 trials, or following patient treatment profiles. The samples of a particular standard  
may be normal or similarly abnormal.

2 The term "stringent hybridization conditions", refers to an overnight  
incubation at 42 °C in a solution comprising 50 % formamide, 5x SSC (750 mM  
5 NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's  
solution, 10 % dextran sulfate and 20 µg/ml denatured sheared salmon sperm DNA,  
followed by washing the filters in 0.1x SSC at about 65 °C. Also contemplated are  
nucleic acid molecules that hybridize to the polynucleotides of the present invention  
at lower stringency hybridization conditions. Changes in the stringency of  
10 hybridization and signal detection are primarily accomplished through the  
manipulation of formamide concentration (lower percentages of formamide result in  
lower stringency); salt conditions, or temperature. For example, lower stringency  
conditions include an overnight incubation at 37 °C in a solution comprising 6x SSPE  
(20X SSPE=3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30 %  
15 formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50 °C  
with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes  
performed following stringent hybridization can be done at higher salt concentrations  
(e.g. 5X SSC). Variation in the above conditions may be accomplished through the  
inclusion and / or substitution of alternative blocking reagents used to suppress  
20 background in hybridization experiments. Typical blocking reagents include  
Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and  
commercially available proprietary formulations. The inclusion of specific blocking  
reagents may require modification of the hybridization conditions described above,  
due to problems with compatibility. A polynucleotide which hybridizes only to  
25 polyA<sup>+</sup> sequences (such as any 3' terminal polyA<sup>+</sup> tract of a cDNA shown in the  
sequence listing), or to a complementary stretch of T (or U) residues would not be  
included in the definition of a "polynucleotide" since such a polynucleotide would  
hybridize to any nucleic acid molecule containing a poly (A) stretch, or the  
complement thereof.

30 The term "target" refers to a biochemical entity involved a biological process.

Targets are typically proteins that play a useful role in the physiology or biology of an

organism. A therapeutic chemical binds to target to alter or modulate its function. As used herein, targets can include cell surface receptors, G-proteins, kinases, ion channels, phospholipases, proteases and other proteins mentioned herein.

The term "test chemical" refers to a chemical to be tested by one or more screening method(s) of the invention as a putative modulator. A test chemical can be any chemical, such as an inorganic chemical, an organic chemical, a protein, a peptide, a carbohydrate, a lipid, or a combination thereof. Usually, various predetermined concentrations of test chemicals are used for screening, such as 0.01 micromolar, 1 micromolar and 10 micromolar. Test chemical controls can include the measurement of a signal in the absence of the test compound or comparison to a compound known to modulate the target.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage identical to a sequence", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two

sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods selected. The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage identical to a sequence" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 30 percent sequence identity, preferably at least 50 to 60 percent sequence identity, more usually at least 60 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 30 percent sequence identity, preferably at least 40 percent sequence identity, more preferably at least 50 percent

sequence identity, and most preferably at least 60 percent sequence identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having  
5 aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and  
10 histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamic-aspartic, and asparagine-glutamine.

Since the list of technical and scientific terms cannot be all encompassing, any  
15 undefined terms shall be construed to have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Furthermore, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "restriction enzyme" or a "high fidelity enzyme" may include mixtures of such enzymes and any other enzymes  
20 fitting the stated criteria, or reference to the method includes reference to one or more methods for obtaining cDNA sequences which will be known to those skilled in the art or will become known to them upon reading this specification.

Before the present sequences, variants, formulations and methods for making and using the invention are described, it is to be understood that the invention is not to  
25 be limited only to the particular sequences, variants, formulations or methods described. The sequences, variants, formulations and methodologies may vary, and the terminology used herein is for the purpose of describing particular embodiments. The terminology and definitions are not intended to be limiting since the scope of protection will ultimately depend upon the claims.

30



## I. MULTIMERIZED DESTABILIZATION DOMAINS

Destabilization domains include proteins, protein domains and amino acid  
5 sequences that when functionally coupled to a target protein effect a change in the  
half-life of that protein when expressed in a cell. Examples include PEST domains,  
stretches of hydrophobic amino acids, phosphorylation dependent degradation signals,  
cyclin destruction boxes and the addition of ubiquitin domains. Preferred as a  
destabilization domain is ubiquitin and homologs thereof, particularly mutants or  
10 homologs comprising mutations that prevent, or significantly reduce, the cleavage of  
ubiquitin multimers by  $\alpha$ -NH-ubiquitin protein endoproteases. In general,  
destabilization domains function by causing the target protein to be recognized by one  
or more elements of the cellular protein degradation apparatus. Once marked for  
destruction, the protein is actively recruited into the 28S proteasome where the protein  
15 is degraded. Within the cell a variety of signals may target a protein for degradation.  
In some cases a destabilization feature may be revealed in a protein as a result of  
oxidation, mis-folding or proteolysis. For example, stretches of hydrophobic amino  
acids are often exposed in denatured or improperly folded proteins thereby targeting  
them for degradation. Short stretches of hydrophobic amino acids, or hydrophobic  
20 domains, also occur in correctly folded proteins and have been identified in proteins  
with short half lives.

For example, the Deg 1 domain of yeast mating type transcription factor  $\alpha 2$  is  
a 19 residue element that forms an amphipathic helix with an exposed hydrophobic  
face, and is responsible for the rapid degradation of this protein (Johnson *et al.*,  
25 (1998) Cell 94 217-227). These elements are believed to be recognized by E3  
ubiquitin ligases and target the protein to degradation through the ubiquitin system  
described below.

PEST domains (regions rich in the amino acids proline (P), glutamic acid (E),  
serine (S) and threonine (T)) are often located at the C-terminal domains of relatively  
30 unstable proteins. (Rogers, *et al.*, (1986) Science 234 (4774) 364-8). A well  
characterized PEST domain is located in residues 422 to 461 of ornithine

decarboxylase, and has been used to successfully destabilize a number of proteins including the green fluorescent protein from *Aequorea* green fluorescent protein (*Li et al. J. Biol. Chem.* (1998) 273 (52) 34970-5). Certain PEST sequences are believed to be recognized by the 26S proteasome subunit directly and do not require ubiquitination.

PEST sequences may also be regulated by phosphorylation, for example multiple phosphorylation within the PEST sequences of the yeast G1 cyclins Cln3 and Cln2 are required for degradation.

Phosphorylation dependent degradation signals have also been identified in the transcription factors NF- $\kappa$ B and  $\beta$ -catenin, in addition to many cell cycle regulatory proteins such as cyclins. (Ghosh *et al.*, (1998) *Ann. Rev. Immunol.* 16 225-260; Aberle *et al.*, (1997) *EMBO J.* 16 3797-3804; Koepp *et al.*, (1999) 97 431-434). These proteins include phosphorylation dependent recognition sequences that bind to one of the growing family of E3 ubiquitin ligases only when the site is phosphorylated. In NF- $\kappa$ B, the binding domain for the E3 ubiquitin ligase comprises the relatively short sequence DS\*GLDS\*, (SEQ. ID. NO.: 1) where S\* denotes phosphoserine. Binding to the E3 ubiquitin ligase does not require a ubiquitination conjugation site in this case.

The cell-cycle destruction box is a partially conserved 9 amino acid sequence motif usually located approximately 40-50 amino acid residues from the N-terminus of the protein first described for the A and B type cyclins. The consensus destruction box sequence has the general structure as shown in **Table 1** below.

TABLE 1								
Consensus destruction box sequence								
R	(A/T)	(A)	L	(G)	x	(I/V)	(G/T)	(N)
1	2	3	4	5	6	7	8	9

Amino acid residues, or combinations of two residues, that appear in parentheses in the above structure occur in more than 50 % of known destruction sequences. The residues at positions 1 and 4 are conserved in all destruction boxes.

Ubiquitin (SEQ. ID. NO.: 2), a 76 amino acid polypeptide found in all eukaryotic cells, is centrally involved in the mechanism of targeting a protein for degradation by the cell. In general, the covalent attachment of a ubiquitin domain (SEQ. ID. NO.: 2), to a protein represents a primary recognition motif for binding of that protein to the proteasome. The attachment of ubiquitin (SEQ. ID. NO.: 2) to the protein typically occurs after recognition of one or more of the destabilization domains discussed above, or some other destabilizing feature of a protein. Attachment of ubiquitin (SEQ. ID. NO.: 2) occurs via the reversible isopeptide linkage of the carboxy-terminus of ubiquitin (SEQ. ID. NO.: 2) to lysine residues in the target protein. After the addition of the first ubiquitin domain (SEQ. ID. NO.: 2), further ubiquitin moieties (SEQ. ID. NO.: 2) may subsequently be added via free lysine residues in ubiquitin (SEQ. ID. NO.: 2) to create branched poly-ubiquitin chains on the substrate protein. These reactions are catalyzed by a family of enzymes that are often referred to as the ubiquitination complex. Once the target protein comprises one or more copies of ubiquitin (SEQ. ID. NO.: 2) it binds with high affinity to the proteasome where it is degraded. (See generally, Hershko *et al.*, (1998) *Annu. Rev. Biochem.* 76 425-79; Laney *et al.*, (1999) *Cell* 97 427-430).

The ubiquitin gene typically comprises multiple copies of the ubiquitin coding sequence (SEQ. ID. NO.: 2). Individual ubiquitin domains (SEQ. ID. NO.: 2) are post-translationally formed from the poly-ubiquitin gene by cleavage of the expressed protein by specific  $\alpha$ -NH-ubiquitin protein endoproteases that are present within all eukaryotic cells. (Jonnalagadda *et al.*, (1989) *J. Biol. Chem.* 264 10637-10642. The endoproteases will cleave either multiple ubiquitin – ubiquitin chains, or ubiquitin - fusion protein constructs, provided that the last amino acid of the ubiquitin moiety (SEQ. ID. NO.: 2) is glycine. If this last amino acid is mutated to a more bulky amino acid the ubiquitin fusion protein is not cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases.

The present inventors have recognized for the first time that the creation of multiple ubiquitin fusion proteins that are not cleavable by the  $\alpha$ -NH-ubiquitin protein endoproteases provides for a facile and tunable method of regulating protein stability.

This invention has many important applications for developing novel assays for intracellular activities, and as a regulatable method of coordinately controlling protein concentrations within the cell.

## 5 II. REPORTER MOIETIES

Enzymatic reporter moieties include any protein capable of catalyzing the creation of a detectable product. Specific examples include alkaline phosphatase,  $\beta$ -galactosidase, chloramphenicol acetyltransferase,  $\beta$ -glucuronidase, peroxidase,  $\beta$ -lactamase, catalytic antibodies, luciferases and other bioluminescent proteins.

Alkaline phosphatase, including human placental and calf intestinal alkaline phosphatase (for example, GenBank Accession # U89937), can be measured using colorimetric, fluorescent and chemiluminescent substrates. (Berger, J., *et al.* (1988) Gene 66 1-10; Kain, S.R. (1997) Methods. Mol. Biol. 63 49-60) Alkaline phosphatase is widely used in transcriptional assays, typically by measuring secreted alkaline phosphatase (SEAP).

$\beta$ -galactosidase ( $\beta$ -Gal) the gene product of the bacterial gene LacZ, is also widely used as a reporter gene for transcriptional analysis and may be assayed via histochemical, fluorescent or chemiluminescent substrates, either within intact, or permeabilized cells. (See, U.S. Patent No. 5,070,012, issued Dec, 3, 1991 to Nolan *et al.*, and Bronstein, I., *et al.*, (1989) J. Chemilum. Biolum. 4 99-111).

$\beta$ -glucuronidase (GUS) is widely used for transcriptional analysis in higher plants and may also be assayed using a variety of histochemical and fluorescent substrates. (See generally U.S. Patent No. 5,599,670, issued Feb, 4, 1997 to Jefferson).

Chloramphenicol acetyltransferase (CAT), encoded by the bacterial Tn9 gene, is widely used for transcriptional assays and is traditionally measured using a radioisotopic assay in cell extracts (See Gorman *et al.*, (1982) 2 1044-51).

Catalytic antibodies are also amenable for use as reporter genes, if the reaction catalyzed by the antibody results in the formation of a detectable product. Examples

include the aldolase specific antibodies 38C2 and 33F12 that catalyze the synthesis of novel fluorogenic retro-aldol reactions (List *et al.*, (1998) Proc. Natl. Acad. Sci USA 95 15351-15355). Typical antibody substrates are cell permeant nonpolar organic molecules that are not substrates for the natural enzymes and are thus good markers of enzyme activity.

### *$\beta$ -Lactamases*

A large number of  $\beta$ -lactamases have been isolated and characterized, all of which would be suitable for use in accordance with the present method. Initially,  $\beta$ -lactamases were divided into different classes (I through V) on the basis of their substrate and inhibitor profiles and their molecular weight (Richmond, M. H. and Sykes, R. B., (1973) Adv. Microb. Physiol. 9 31-88). More recently, a classification system based on amino acid and nucleotide sequence has been introduced (Ambler, R.P., (1980) Phil. Trans. R. Soc. Lond. [Ser.B.] 289 321-331). Class A  $\beta$ -lactamases possess a serine in the active site and have an approximate weight of 29kd. This class contains the plasmid-mediated TEM  $\beta$ -lactamases such as the RTEM enzyme of pBR322. Class B  $\beta$ -lactamases have an active-site zinc bound to a cysteine residue. Class C enzymes have an active site serine and a molecular weight of approximately 39kd, but have no amino acid homology to the class A enzymes.

The coding regions of an exemplary  $\beta$ -lactamase employed in the methods described herein include SEQ. ID. NOs: 3 through 7. Nucleic acids encoding proteins with  $\beta$ -lactamase activity can be obtained by methods known in the art, for example, by polymerase chain reaction of cDNA using primers based on a DNA sequence in SEQ. ID. NO.: 3. PCR methods are described in, for example, U.S. Patent No. 4,683,195; Mullis *et al.* (1987) Cold Spring Harbor Symp. Quant. Biol. 51 263; and Erlich, ed., PCR Technology, (Stockton Press, NY, 1989).

Preferably, beta-lactamase polynucleotides encode an intracellular form of a protein with beta-lactamase activity that lacks a functional signal sequence. This provides the advantage of trapping the normally secreted beta-lactamase protein within the cell, which enhances the signal to noise ratio of the signal associated with beta-lactamase activity, and enables the individual cells to be FACST<sup>TM</sup> sorted. For

example, in any of the polypeptides of SEQ. ID. NO.: 3-7, the signal sequence has been replaced with the amino acids Met-Ser. Accordingly, upon expression, beta-lactamase activity remains within the cell. For expression in mammalian cells it is preferable to use beta-lactamase polynucleotides with nucleotide sequences preferred by mammalian cells. In some applications secreted forms of beta-lactamase can be used with the methods described herein.

A variety of colorimetric and fluorescent substrates of  $\beta$ -lactamase are available. Fluorescent substrates include those capable of changes, either individually or in combination, of total fluorescence, excitation or emission spectra or fluorescence resonance energy transfer (FRET), for example those described in U.S. Patent No. 5,741,657, issued April 21, 1998, and U.S. Patent 5,955,604, issued September 22, 1999. Any membrane permanent  $\beta$ -lactamase substrate capable of being measured inside the cell after cleavage can be used in the methods and compositions of the invention. Membrane permanent  $\beta$ -lactamase substrates will not require permeablizing eukaryotic cells either by hypotonic shock or by electroporation. Generally, such non-specific pore forming methods are not desirable to use in eukaryotic cells because such methods injure the cells, thereby decreasing viability and introducing additional variables into the screening assay (such as loss of ionic and biological contents of the shocked or electropolated cells). Such methods can be used in cells with cell walls or membranes that significantly prevent or retard the diffusion of such substrates. Preferably, the membrane permeant  $\beta$ -lactamase substrates are transformed in the cell into a  $\beta$ -lactamase substrate of reduced membrane permeability (usually at least five-fold less permeable) or that is membrane impermeant. Transformation inside the cell can occur via intracellular enzymes (e.g. esterases) or intracellular metabolites or organic molecules (e.g. sulfhydryl groups).

### *Bioluminescent proteins*

Preferred bioluminescent proteins include firefly, bacterial or click beetle luciferases, aequorins and other photoproteins, for example as described in U.S. Patents 5,221,623, issued June 22, 1989 to Thompson *et al.*, U.S. Patent No. 5,683,888 issued November 4, 1997 to Campbell; U.S. Patent No. 5,674,713 issued

September 7 1997 to DeLuca *et al.*, U.S. Patent No. 5,650,289 issued July 22, 1997 to Wood and U.S. Patent No. 5,843,746 issued December 1, 1998 to Tatsumi *et al.* Particularly preferred are bioluminescent proteins isolated from the ostracod *Cypridina* (or *Vargula*) *hilgendorffii*. (Johnson and Shimomura, (1978) Methods  
5 Enzymol 57 331-364; Thompson, Nagata & Tsuji (1989) Proc. Natl. Acad. Sci. USA 86, 6567-6571).

Beyond the availability of bioluminescent proteins (luciferases) isolated directly from the light organs of beetles, cDNAs encoding luciferases of several beetle species (including, among others, the luciferase of *P. pyralis* (firefly), the four  
10 luciferase isozymes of *P. plagiophthalmus* (click beetle), the luciferase of *L. cruciata*(firefly) and the luciferase of *L. lateralis*) (deWet *et al.*, (1987) Molec. Cell. Biol. 7, 725-737; Masuda *et al.*, (1989) Gene 77, 265-270; Wood *et al.*, (1989) Science 244, 700-702; European Patent Application Publication No. 0 353 464) are available. Further, the cDNAs encoding luciferases of any other beetle species, which  
15 make bioluminescent proteins, are readily obtainable by the skilled using known techniques (de Wet *et al.* (1986) Meth. Enzymol. 133, 3- 14; Wood *et al.*, (1989) Science 244, 700-702).

Most firefly and click beetle luciferases are ATP- and magnesium dependent and require oxygen for light production. Typically light emission from these enzymes  
20 exhibits a rapid burst in intensity followed by a rapid decrease in the first few seconds, followed by a significantly slower sustained light emission. Relatively sustained light output at high rates has been accomplished in these systems by inclusion of coenzyme A, dithiothreitol and other reducing agents that reduce product inhibition and slows inactivation of the luciferase that occurs during catalysis of the  
25 light producing reaction, as described in U.S. Patents No. 5,641,641, issued June 24, 1997, and U.S. Patent No. 5,650,289, issued July 22, 1997. Such stable light emitting systems are preferred for use in the present invention.

Particularly preferred bioluminescent proteins are those derived from the ostracod *Cypridina* (or *Vargula*) *hilgendorffii*. The *Cypridina* luciferase (GenBank  
30 accession no. U89490) uses no cofactors other than water and oxygen, and its luminescent reaction proceeds optimally at pH 7.2 and physiological salt

concentrations, (Shimomura, O., Johnson, F.H. and Saiga, Y. (1961) J. Cell. Comp. Physiol. 58 113-124). By comparison, firefly luciferase has optimal activity at low ionic strength, alkaline pH and reducing conditions, that are typically quite different to those usually found within mammalian cells. Because *Cypridina* luciferase has a turnover number of 1600 min<sup>-1</sup> and a quantum yield of 0.29, (Shimomura, O. & Johnson, F.H. and Masugi, T. (1969) Science 164 1299-1300; Shimomura, O. & Johnson, F.H. (1970) Photochem. Photobiol. 12 291-295), the *Cypridina* luciferase produces a specific photon flux exceeding that of the optimized firefly system by a factor of at least 50 (Miesenbock and Rothman, (1997) Proc. Natl. Acad. Sci. USA 94 3402-3407).

### *Naturally Fluorescent Proteins*

Another preferred class of embodiments of the reporter moiety includes naturally fluorescent proteins such as the Green Fluorescent Protein (GFP) of *Aequorea victoria* (Tsien, R.Y. (1998) Annu. Rev. Biochem. 67 509-44). Because the entire fluorophore and peptide of a naturally fluorescent protein can be expressed within intact living cells without the addition of other co-factors or fluorophores, optical probes comprising such proteins as the reporter moiety provide the ability to monitor activities, within defined cell populations, tissues or in an entire transgenic organism. For example, by the use of cell type specific promoters and subcellular targeting motifs, it is possible to selectively target the probe to a discrete location to enable highly spatially defined measurements.

Naturally fluorescent proteins have been isolated and cloned from a number of marine species including the sea pansies *Renilla reniformis*, *R. kollikeri* and *R. mullerei* and from the sea pens *Ptilosarcus*, *Stylatula* and *Acanthoptilum*, as well as from the Pacific Northwest jellyfish, *Aequorea victoria*; Szent-Gyorgyi *et al.* (SPIE conference 1999); D.C. Prasher *et al.*, (1992) Gene, 111:229-233 and several species of coral (Matz *et al.* (1999). Nature Biotechnology 17 969-973. These proteins are capable of forming a highly fluorescent, intrinsic chromophore through the cyclization and oxidation of internal amino acids within the protein that can be



spectrally resolved from weakly fluorescent amino acids such as tryptophan and tyrosine.

Additionally naturally fluorescent proteins have also been observed in other organisms, although in most cases these require the addition of some exogenous factor to enable fluorescence development. For example, the cloning and expression of yellow fluorescent protein from *Vibrio fischeri* strain Y-1 has been described by T.O. Baldwin *et al.*, Biochemistry (1990) 29 5509-15. This protein requires flavins as fluorescent co-factors. The cloning of Peridinin-chlorophyll *a* binding protein from the dinoflagellate *Symbiodinium* sp. was described by B.J. Morris *et al.*, (1994) Plant Molecular Biology, 24 673:77. One useful aspect of this protein is that it fluoresces in red. The cloning of phycobiliproteins from marine cyanobacteria such as *Synechococcus*, e.g., phycoerythrin and phycocyanin, is described in S.M. Wilbanks *et al.*, (1993) J. Biol. Chem. 268 1226-35. These proteins require phycobilins as fluorescent co-factors, whose insertion into the proteins involves auxiliary enzymes. The proteins fluoresce at yellow to red wavelengths.

A variety of mutants of the GFP from *Aequorea victoria* have been created that have distinct spectral properties, improved brightness and enhanced expression and folding in mammalian cells compared to the native GFP, (SEQ. ID. NO.: 8), **Table 2.** (*Green Fluorescent Proteins*, Chapter 2, pages 19 to 47, edited Sullivan and Kay, Academic Press, U.S. Patent Nos: 5,625,048 to Tsien *et al.*, issued April 29, 1997; 5,777,079 to Tsien *et al.*, issued July 7, 1998; and U.S. Patent No. 5,804,387 to Cormack *et al.*, issued September 8, 1998). In many cases these functional engineered fluorescent proteins have superior spectral properties to wild-type *Aequorea* GFP and are preferred for use as reporter moieties in the present invention.

TABLE 2					
Aequorea Fluorescent Proteins					
Mutations	Common Name	Quantum Yield (Φ) & Molar Extinction (ε)	Excitation & Emission Max	Relative Fluorescence At 37 °C	Sensitivity To Low pH % max F at pH 6
S65T type					

S65T, S72A, N149K, M153T, I167T	Emerald (SEQ. ID. NO.: 28)	$\Phi = 0.68$ $\epsilon = 57,500$	487 509	100	91
F64L, S65T, V163A		$\Phi = 0.58$ $\epsilon = 42,000$	488 511	54	43
F64L, S65T	EGFP	$\Phi = 0.60$ $\epsilon = 55,900$	488 507	20	57
S65T		$\Phi = 0.64$ $\epsilon = 52,000$	489 511	12	56
<b>Y66H type</b>					
F64L, Y66H, Y145F, V163A	P4-3E	$\Phi = 0.27$ $\epsilon = 22,000$	384 448	100	N.D.
F64L, Y66H, Y145F		$\Phi = 0.26$ $\epsilon = 26,300$	383 447	82	57
Y66H, Y145F	P4-3	$\Phi = 0.3$ $\epsilon = 22,300$	382 446	51	64
Y66H	BFP	$\Phi = 0.24$ $\epsilon = 21,000$	384 448	15	59
<b>Y66W type</b>					
S65A, Y66W, S72A, N146I, M153T, V163A	W1C	$\Phi = 0.39$ $\epsilon = 21,200$	435 495	100	82
F64L, S65T, Y66W, N146I, M153T, V163A	W1B	$\Phi = 0.4$ $\epsilon = 32,500$	434 452 476 (505)	80	71
Y66W, N146I, M153T, V163A	hW7	$\Phi = 0.42$ $\epsilon = 23,900$	434 452 476 (505)	61	88
Y66W			436 485	N.D.	N.D.
<b>T203Y type</b>					
S65G, S72A, K79R, T203Y	Topaz	$\Phi = 0.60$ $\epsilon = 94,500$	514 527	100	14
S65G, V68L, S72A, T203Y	10C	$\Phi = 0.61$ $\epsilon = 83,400$	514 527	58	21
S65G, V68L, Q69K, S72A, T203Y	h10C+	$\Phi = 0.71$ $\epsilon = 62,000$	516 529	50	54
S65G, S72A, T203H		$\Phi = 0.78$ $\epsilon = 48,500$	508 518	12	30
S65G, S72A, T203F		$\Phi = 0.70$ $\epsilon = 65,500$	512 522	6	28
<b>T203I type</b>					
T203I, S72A, Y145F	Sapphire	$\Phi = 0.64$ $\epsilon = 29,000$	395 511	100	90
T203I, T202F	H9	$\Phi = 0.6$ $\epsilon = 20,000$	395 511	13	80

Non Aequorea, naturally fluorescent proteins, for example Anthozoan fluorescent proteins, and functional engineered homologs thereof, are also suitable for use in the present invention including those shown in **Table 3** below.

<b>TABLE 3</b> <b>Anthozoa Fluorescent Proteins</b>					
Species	Protein Name	Quantum Yield ( $\Phi$ ) & Molar Extinction ( $\epsilon$ )	Excitation & Emission Max	Relative Brightness	SEQ. ID. NO.:
<i>Anemonia majano</i>	amFP486	$\Phi = 0.24$ $\epsilon = 40,000$	458 486	0.43	SEQ. ID. NO.: 9
<i>Zoanthus sp</i>	zFP506	$\Phi = 0.63$ $\epsilon = 35,600$	496, 506	1.02	SEQ. ID. NO.:10
	zFP538	$\Phi = 0.42$ $\epsilon = 20,200$	528, 538	0.38	SEQ. ID. NO.:11
<i>Discosoma striata</i>	dsFP483	$\Phi = 0.46$ $\epsilon = 23,900$	443 483	0.5	SEQ. ID. NO.:12
<i>Discosoma sp</i> "red"	drFP583	$\Phi = 0.23$ $\epsilon = 22,500$	558 583	0.24	SEQ. ID. NO.:13
<i>Clavularia sp</i>	CFP484	$\Phi = 0.48$ $\epsilon = 35,300$	456 484	0.77	SEQ. ID. NO.:14

5

### III. LINKER MOIETIES

Generally linker moieties for measuring a post-translational activity encompass a post-translational recognition motif that contains a residue that, when modified, modulates the coupling of the reporter moiety to the multimerized destabilization domain, thus effecting a change in the stability of the reporter moiety. Typically, for measuring proteases, such linkers contain a single scissile bond (bond that is cleaved within the substrate) for a specific protease and preserve the native function and activity of the reporter moiety and destabilization domains in the intact fusion protein. The design and size of peptide sequences for specific constructs, is dependent upon the application for which the optical probe is to be used. For example, for most applications, the peptide linker separating the reporter moiety and

10

15

the multimerized destabilization domains will typically be in the range of 5 to 50 amino acids in length, preferably 10 to 25 amino acids in length, or more preferably 10 to 15 amino acids in length. For certain applications, the peptide may be significantly larger, up to and including entire protein domains, for example 50 to 100 amino acids in length. Smaller peptides, in the range of 5 to 50 amino acids may also be used. Typically the protease site may be located at any position within the linker with respect to the reporter moiety and destabilization domains.

In one embodiment the linker comprises a single polypeptide chain that covalently couples the destabilization domains to the reporter moiety. Typically in this embodiment, the linker will comprise a post-translational recognition motif such as a protease recognition motif. Cleavage of the linker by the protease at the cleavage site results in uncoupling of the multimerized destabilization domains from the reporter moiety resulting in a modulation in the stability of the reporter moiety. An important feature of the linker is that it does not contain a protease recognition site for  $\alpha$ -NH-ubiquitin protein endoproteases that would otherwise result in the post-translational processing of the construct irrespective of the presence or absence of the target post-translational activity. Any cleavage activity capable of hydrolyzing the linker moiety may be assayed with this embodiment of the present invention, provided it does not also cleave the reporter moiety thereby directly modulating its function.

In another aspect of this method, the linker may comprise distinct post-translational recognition motifs and cleavage sites for example, a phosphorylation site and a protease cleavage site, as described in commonly owned U.S. Patent Application No. 09/306,542 filed May 5, 1999. In this case, post-translational modification of the linker results in the modulation of the rate and efficiency of cleavage of the modified linker compared to the non-modified linker. This approach enables the present method to be used to detect a broad range of post translational activities.

In some embodiments, the linker functions to couple a target protein to one or more destabilization domains for the purpose of regulating the concentration of the target protein in the cell. In this case the linker need not contain a protease cleavage

site, and may be significantly smaller, in the order of about 1 to 10 amino acids in length.

In another aspect, the linker may comprise two separate polypeptide chains that are capable of interacting with each other to functionally couple the multimerized destabilization domains to the reporter gene. This approach enables an additional range of post-translational activities to be assayed. In this embodiment, one polypeptide chain is typically covalently coupled to the multimerized destabilization domain, and a separate polypeptide chain is covalently coupled to the reporter moiety. (FIG. 1) Binding of the first polypeptide chain to the second polypeptide chain results in coupling of the destabilization domain to reporter moiety resulting in a modulation of the stability of the reporter moiety. This approach thus enables the identification and detection of protein-protein interactions between defined proteins as well as the ability to detect post-translational modifications that influence these protein-protein interactions.

Examples of suitable interaction domains include protein-protein interaction domains such as SH2, SH3, PDZ, 14-3-3, WW and PTB domains. Other interaction domains are described in for example, the database of interacting proteins available on the web at <http://www.doe-mbi.ucla.edu>.

To identify and characterize the interaction of two test proteins, the method would typically involve 1) the creation of a first fusion protein comprising the first test protein coupled to the reporter moiety, and a second fusion protein comprising the second test protein coupled to the multimerized destabilization domain construct. 2) The introduction of the test protein fusion proteins alone in to control cells, and in combination into test cells. 3) The measurement of the stability of the reporter moiety in the control cells and test cells. 4) Comparison of the stability of the reporter moiety in the control cells, compared to the stability of the reporter moiety in the test cells. If the cell expressing both test fusion proteins exhibits a reporter moiety with a significantly altered stability (or level of expression) compared to the control cells, then the results indicate that the two proteins do interact under the experimental conditions chosen. Conversely if the stability's of the reporter moieties in the control

cells, and in the test cells are the same, then the results indicate that the proteins probably don't interact strongly under the test conditions.

The method also enables the detection and characterization of stimuli (such as receptor stimulation) that cause two proteins to alter their degree of interaction. In this case, a cell line is created that expresses the first and second fusion proteins, as described above, comprising interaction domains that exhibit, or are believed to exhibit post-translational regulated interactions. For example, post-translational modification by phosphorylation of serine or threonine residues can modulate 14-3-3 domain interactions, tyrosine phosphorylation can influence SH2 domain interactions, the redox state can influence disulfide bond formation. The cell line is then exposed to a test stimulus to determine whether the stimulus regulates the interaction of the two proteins. If the stimulus does regulate the interaction of the two proteins, then this will result in the coupling of the multimerized destabilization domain fusion protein to the reporter moiety fusion protein, subsequently resulting in a modulation of the stability of the reporter moiety in the treated cells, compared to the non-treated cells.

The invention is also readily amenable to identifying new protein-protein interactions. For example, where a first protein is known, but the protein(s) with which it interacts are unknown. In this case, a first fusion protein is made between the first protein and the reporter moiety (or destabilization domain) and cloned into a suitable expression vector. Second, a library of test proteins, for example isolated from a cDNA expression library, is fused in frame to the multimerized destabilization domains (or reporter moiety) and subcloned into a second expression vector. Typically the first fusion protein would be then be introduced into a population of test cells and single clones identified that stably expressed the reporter moiety. The library of test proteins (typically in the form of expression vectors) would be introduced into the clonal cells, stably expressing the first fusion protein. The resulting transformed cells would then be screened to identify cells with altered expression of the reporter moiety fusion compared to the control cells. Suitable clones expressing the reporter moieties with modulated stability, (i.e., reduced levels of the reporter moiety) may then be identified, isolated and characterized, for example by fluorescence activated

cell sorting (FACS™ ). Those library members that display reporter moieties with larger relative changes in expression level may then be identified by the degree to which the stability of the reporter moiety is altered for each library member after exposure to the library of test fusion proteins.

5

#### IV. METHODS OF USE

##### *Introduction of constructs into cells*

Typically the constructs of the present invention will be introduced and  
10 expressed in target cells via the use of standard molecular biology techniques known in the art. Another approach involves the use of membrane translocating sequences, as described in U.S. patent No. 5,807,746, issued Sept 15 1998 to Lin *et al.* to introduce the protein constructs into cells.

Nucleic acids may also be used to transfect cells with sequences coding for  
15 expression of the multimerized destabilization domain, linker and reporter moiety. Generally these will be in the form of an expression vector including expression control sequences operatively linked to a nucleotide sequence coding for expression of the polypeptide. As used, the term "nucleotide sequence coding for expression of" a polypeptide refers to a sequence that, upon transcription and translation of mRNA,  
20 produces the polypeptide. This can include sequences containing, *e.g.*, introns. As used herein, the term "expression control sequences" refers to nucleic acid sequences that regulate the expression of a nucleic acid sequence to which it is operatively linked. Expression control sequences are operatively linked to a nucleic acid sequence when the expression control sequences control and regulate the transcription  
25 and, as appropriate, translation of the nucleic acid sequence. Thus, expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (*i.e.*, ATG) in front of a protein-encoding gene, splicing signals for introns, IRES sequences (internal ribosome entry site) maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop  
30 codons.

Methods that are well known to those skilled in the art can be used to construct expression vectors containing the multimerized destabilization domain, linker, reporter moiety construct. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination.

5 (See, for example, the techniques described in Maniatis, *et al.*, (1989) Cold Spring Harbor Laboratory, N.Y.). Many commercially available expression vectors are available from a variety of sources including Clontech (Palo Alto, CA), Stratagene (San Diego, CA) and Invitrogen (San Diego, CA) as well as and many other commercial sources.

10 A contemplated version of the method is to use inducible controlling nucleotide sequences to produce a sudden increase in the expression of the reporter moiety, linker and multimerized destabilization domain construct e.g., by inducing expression of the construct. Example inducible systems include the tetracycline inducible system first described by Bujard and colleagues (Gossen and Bujard (1992)  
15 Proc. Natl. Acad. Sci USA 89 5547-5551, Gossen *et al.* (1995) Science 268 1766-1769) and described in U.S. Patent No 5,464,758.

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as *E. coli*, competent cells that are capable of DNA uptake  
20 can be prepared from cells harvested after exponential growth phase and subsequently treated by the CaCl<sub>2</sub> method by procedures well known in the art. Alternatively, MgCl<sub>2</sub> or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as  
25 calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be co-transfected with DNA sequences encoding the fusion polypeptide of the invention, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex  
30 thymidine kinase gene. Another method is to use an eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform



eukaryotic cells and express the protein. (*Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Preferably, an eukaryotic host is utilized as the host cell as described herein.

5 The construction of expression vectors and the expression of genes in transfected cells involve the use of molecular cloning techniques also well known in the art. Sambrook *et al.*, (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (most recent Supplement). Nucleic acids used to transfect cells with  
10 sequences coding for expression of the polypeptide of interest generally will be in the form of an expression vector including expression control sequences operatively linked to a nucleotide sequence coding for expression of the polypeptide comprising the optical probe.

#### 15 *Assays for post-translational activities*

In one class of embodiments, the present invention can be used to measure post-translational activities, such as proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation, disulfide bond formation and ADP-ribosylation within cells.

20 The method generally involves the expression within, or introduction into a cell of a reporter moiety that is functionally coupled to one or more destabilizing domains via a linker. The linker typically contains a recognition motif that is specific for the post-translational activity to be assayed. Modification of the linker by the post-translational activity, results in uncoupling of the reporter moiety from the destabilizing domain resulting in a modulation in the stability of the reporter moiety.  
25 The level of activity within a sample is sensed by a measurable change in the level of the reporter moiety, for example by detecting at least one optical property of the reporter moiety, or by detecting at least one optical property of detectable product of the reporter moiety.

30 To measure protease activity, it is typically desirable to provide an expression vector in which the expressed fusion gene product comprises a reporter moiety

covalently linked to the multimerized destabilization moieties via a single amino acid chain. Thus under these conditions the expressed construct is destabilized until acted upon by the target protease. Upon proteolysis, the cleaved reporter moiety exhibits significantly increased stability, resulting in its steady state accumulation within the cell to a higher level.

The choice of reporter moiety depends in part on the cellular system in which the assays are conducted, and the sensitivity and detection means at hand. For mammalian cells, the  $\beta$ -lactamase,  $\beta$ -galactosidase, and naturally fluorescent protein based reporter genes provide for intracellular fluorescent measurements, which are preferred. Preferred reporter moieties for luminescent readouts include luciferase and other bioluminescent protein based reporters. In plant studies, preferred reporters include  $\beta$ -glucuronidase and luciferase. For transgenic applications in whole animals or intact tissue samples, naturally fluorescent proteins are preferred because the reporter does not require the addition of any substrates or co-factors in order to produce a detectable product. For applications where high sensitivity is required, for example because the target activity has a low turnover number, enzymatic reporter moieties are preferred because they provide enzymatic amplification. That is, each reporter moiety is capable of generating hundreds or thousands of detectable products per minute. By comparison a non enzymatic reporter, such as a naturally fluorescent protein, provides for little signal amplification.

The choice of the multimerized destabilization domain, and the number of copies of the destabilization domain to use are also dependent on the reporter moiety and type of activity being measured. Preferred destabilization domains include, those based on ubiquitin (SEQ. ID. NO.: 2) and mutants and homologs thereof. Particularly preferred are mutants or homologs of ubiquitin (SEQ. ID. NO.: 2) comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by  $\alpha$ -NH-ubiquitin protein endoproteases.

To establish the optimal number of destabilization domains one would generally start by evaluating a construct containing three copies of the destabilization domain. Depending upon the results, one would either increase or decrease the

number of copies of destabilization domains. Generally one would increase the number of copies of the destabilization domain if the steady state levels of the non-protease treated samples were too high (too little degradation), and decrease the number of copies of the destabilization domain if the steady state level of the non-protease treated samples were too low (too much degradation). If the target protein was subject to excessive degradation, the steady state level of the target protein may be too low to provide for effective cleavage by the protease, particularly if that protease exhibits a relatively low affinity for that substrate.

An important advantage of the present invention is the ability to titrate the degree of destabilization, and therefore the steady state concentration, of the target protein in the cell. Since the destabilized, unmodified sensor represents the substrate for the target activity, it is preferable to provide the substrate at a physiologically relevant concentration within the cell while retaining the appropriate turnover characteristics for each individual reporter molecule.

For assays measuring protease activity, the linker generally comprises a protease recognition motif within its sequence. The protease recognition motif may be placed anywhere within the linker moiety, but is conveniently placed close to the center of the linker unless there are steric, or other reasons, to position the recognition motif at a specific location. Typically, the recognition motif will provide for relatively specific recognition of the sequence by the target protease. In some cases it may be preferable for the linker to contain a second "control" protease site for a known protease for use as a positive control.

The expression vector will normally direct expression of the sensor to the cytosol of the cell, although other cellular compartments, such as the plasma membrane are also practical. Once the expression vector is introduced in a population of cells, the cells are typically screened for reporter moiety expression level in the absence of the target protease. This can be achieved by FACS™, after addition of appropriate substrates for the reporter moieties (if required). While cells may be selected for varying levels of expression of the reporter moiety within the population of cells, observations to date suggest that cells exhibiting somewhat lower levels of reporter moiety are superior to those that initially exhibit high levels of reporter

moieties under these conditions. Cells may also be selected via antibiotic resistance to provide for stable cell lines.

Once isolated and characterized, the resulting cell line represents a living sensor for the activation or expression of the target protease that enables the identification and screening of compounds that modulate the activation of the target protease. Importantly these determinations can be completed within the living cell where other issues such as membrane permeability, specificity and toxicity may be directly assessed.

In most cases, it will be preferable to start with a cell line that does not normally express high levels of the active target protease. However if this is not possible, then the initial evaluation of the cell lines may be modified in order to screen for cells initially exhibiting high levels of reporter moiety expression. For example, by using an inhibitor of the reporter moiety to inhibit basal reporter gene activity, (as discussed below). In general any types of cells may be used with the present invention, including animal, plant, insect, yeast and other eukaryotic cells or prokaryotic cells.

In whole cell studies it may be desirable to add an inhibitor of protein synthesis such as cycloheximide in order to reduce the steady state level of the destabilized reporter moiety in the cell immediately prior to the measurement of reporter activity. This approach has the advantage of improving the dynamic range of the assay because in the absence of new protein synthesis, uncleaved and therefore destabilized reporter moieties are destroyed by targeting to the proteasome leaving the cleaved and stabilized reporters intact within the cell. (i.e. the background is reduced). This results in a larger net difference in reporter moiety activity in cells containing a suitable protease compared with those lacking a suitable protease. Typically for such uses, cycloheximide is added to cell in the range of 10 to 150  $\mu\text{g/ml}$  cycloheximide, preferably 50 to 100  $\mu\text{g/ml}$ . Generally cells are pretreated with an appropriate stimulus to activate the target protease, and then cycloheximide is added one to two hours prior to the addition of suitable substrates for the reporter moiety.

In another aspect of this method, it sometimes may also be desirable to add an inhibitor of the enzymatic reporter moiety to reduce the activity of the reporter moiety

prior to compound addition in screening applications. For example, in order to screen for inhibitors of a constitutively active protease, such inhibitors of reporter activity can be used to eliminate the pool of cleaved and stabilized reporter prior to adding compound, in effect zeroing out the cells to begin the experiment. This approach also  
5 has the advantage that the actual concentration of destabilized substrate molecules is not reduced in the cell, so that the protein substrate can be effectively degraded by the target protease. Example inhibitors include clavulanic acid for the  $\beta$ -lactamase reporter gene (see commonly owned U.S. Patent Application No. 09/067,612 filed April 28, 1999) and phenylethyl- $\beta$ -D-thiogalactoside for  $\beta$ -galactosidase (see Fiering  
10 *et al.*, (1991) *Cytometry* 12 291-301). These membrane permeable inhibitors may be added prior to, simultaneously with, or after exposure of the cells to an inhibitor of protein synthesis.

To measure the degree of protein-protein interaction between two defined test proteins, it is typically desirable to separately couple one protein to one or more  
15 destabilization domains, and the second protein to the reporter moiety, and then express both fusion proteins in a test cell. This could be achieved for example by transfecting a cell with two compatible expression vectors. In one expression vector, the expressed fusion protein typically comprises a reporter moiety coupled to the first test protein, and in the second expression vector, the expressed fusion protein  
20 typically comprises the second test protein, coupled to one or more destabilization domains.

If the first polypeptide fusion protein binds to the second polypeptide fusion protein then the destabilization domain(s) are effectively coupled to the reporter moiety resulting in a modulation of its stability. Thus the relative degree of  
25 destabilization of the reporter moiety is a direct indicator of the extent to which the proteins physically interact. Typically this can be accomplished by determining the stability of the reporter moiety in a cell expressing both proteins compared to a control cell, expressing the reporter moiety fusion protein alone. If the cell expressing both constructs exhibits a reporter moiety with a significantly altered stability  
30 compared to the control cell, the results indicate that the two proteins are interacting when co-expressed within the cell.

The choice and selection of the appropriate reporter moiety and destabilization domain are determined by the same issues of sensitivity and ease of detection discussed above. Preferred reporter moieties include  $\beta$ -lactamase and naturally fluorescent proteins. Preferred destabilization domains include those based on ubiquitin (SEQ. ID. NO.: 2), and mutants and functional homologs thereof. Particularly preferred are mutants or homologs of ubiquitin (SEQ. ID. NO.: 2) comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by  $\alpha$ -NH-ubiquitin protein endoproteases.

The choice of the number of copies of the destabilization domain is dependent on the affinity of the target interaction to be measured, and the degree of destabilization exerted on the reporter moiety when the proteins are associated. In many cases, the affinity of the interaction will not be known and it will be necessary to evaluate a range of multimerized constructs in order to identify the optimal assay characteristics. Ideally a multimerized construct will be selected in which both the first test protein and the second test protein are present at physiologically relevant concentrations. One way to achieve this result may be to couple both the first test protein and the second test protein with at least one ubiquitin (SEQ. ID. NO.: 2) domain. Under these circumstances both proteins are slowly degraded when separated, but more rapidly degraded when complexed together.

#### *Induction and regulation of expression levels of target proteins*

In another embodiment, the invention provides for a generalized way of coordinately regulating the cellular concentration of a plurality of target proteins in a cell, or transgenic organism. In this method, the target proteins are operatively coupled to a multimerized destabilization domain via a linker. By varying the number of destabilization domains present in the multimerized destabilization domain, it is possible to titrate the degree of destabilization, and therefore the steady state concentration of the target protein within the cell or transgenic organism. Thus using this approach it is possible to reproducibly vary the relative stoichiometry, as well as, the level of expression, of one or more target proteins.

In some embodiments the linker may comprise about 1 to 10 amino acids. Typically the linker is non-cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases.

In one embodiment the linker may contain a non-naturally occurring protease cleavage site (in that cell type), such that cleavage of the linker by the protease results in uncoupling of the target protein from the multimerized destabilization domain hence creating an increase in the stability and concentration of the target protein after protease digestion. In one aspect of this method, regulation of the activity of the protease can be achieved via regulating the concentration and exposure of the cell to an inhibitor of the protease.

This approach enables the coordinate regulation of the intracellular concentration of a number of target proteins that contain the same protease recognition sites in their linker moieties, simultaneously within a cell. The approach is particularly well suited for the engineering of organisms or cells where multiple proteins need to be induced and expressed in order to create the desired effect, for example for regulating a multi-step metabolic or signal transduction pathway.

In one embodiment the protease is a non-naturally occurring protease in the host cell, which recognizes a relatively rare recognition motif in the linker moiety, for example, including proteases such as Factor Xa (EC 3.4.21.6), Entrokinase (EC 3.4.21.9) and IgA protease (EC 3.4.21.72). Proteases that recognize defined sequences of at least 4, or preferably at least 5 or more preferably about 6 amino acid residues, are generally preferred. Viral proteases, such as a CMV protease or other non-naturally occurring proteases (for that particular cell or organism) are also preferred. If this is the case, then expression of the protease should not significantly impact the cell, and the fusion proteins should not suffer non-specific degradation via the host cells endogenous proteases. Induction or activation of the protease in the cell results in a rapid increase in protease activity within the cell that can cleave the target fusion proteins thereby increasing their stability and steady state concentration in the cell.

## V. SCREENING APPLICATIONS

The present invention is suited for use with systems and methods that utilize automated and integratable workstations for identifying modulators, and chemicals having useful activity. Such systems are described generally in the art (see, U.S. Patent NOs: 4,000,976 to Kramer *et al.* (issued January 4, 1977), 5,104,621 to Pfoest *et al.* (issued April 14, 1992), 5,125,748 to Bjornson *et al.* (issued June 30, 1992), 5,139,744 to Kowalski (issued August 18, 1992), 5,206,568 Bjornson *et al.* (issued April 27, 1993), 5,350,564 to Mazza *et al.* (September 27, 1994), 5,589,351 to Harootunian (issued December 31, 1996), and PCT Application Nos: WO 93/20612 to Baxter Deutschland GMBH (published October 14, 1993), WO 96/05488 to McNeil *et al.* (published February 22, 1996), WO 93/13423 to Agong *et al.* (published July 8, 1993) and U.S. Patent No. 5,985,214, issued November 16, 1999.

Typically, such a system includes: A) a storage and retrieval module comprising storage locations for storing a plurality of chemicals in solution in addressable chemical wells, a chemical well retriever and having programmable selection and retrieval of the addressable chemical wells and having a storage capacity for at least 100,000 addressable wells, B) a sample distribution module comprising a liquid handler to aspirate or dispense solutions from selected addressable chemical wells, the chemical distribution module having programmable selection of, and aspiration from, the selected addressable chemical wells and programmable dispensation into selected addressable sample wells (including dispensation into arrays of addressable wells with different densities of addressable wells per centimeter squared) or at locations, preferably pre-selected, on a plate, C) a sample transporter to transport the selected addressable chemical wells to the sample distribution module and optionally having programmable control of transport of the selected addressable chemical wells or locations on a plate (including adaptive routing and parallel processing), D) a reaction module comprising either a reagent dispenser to dispense reagents into the selected addressable sample wells or locations on a plate or a fluorescent detector to detect chemical reactions in the selected addressable sample wells or locations on a plate, and a data processing and integration module.

The storage and retrieval module, the sample distribution module, and the reaction module are integrated and programmably controlled by the data processing



and integration module. The storage and retrieval module, the sample distribution module, the sample transporter, the reaction module and the data processing and integration module are operably linked to facilitate rapid processing of the addressable sample wells or locations on a plate. Typically, devices of the invention

5 can process at least 100,000 addressable wells or locations on a plate in 24 hours. This type of system is described in commonly owned U.S. Patent No. 5,985,214, issued November 16, 1999. If desired, each separate module is integrated and programmably controlled to facilitate the rapid processing of liquid samples, as well as being operably linked to facilitate the rapid processing of liquid samples. In one

10 embodiment the system provides for a reaction module that is a fluorescence detector to monitor fluorescence. The fluorescence detector is integrated to other workstations with the data processing and integration module and operably linked with the sample transporter. Preferably, the fluorescence detector is of the type described herein and can be used for epi-fluorescence. Other fluorescence detectors that are compatible

15 with the data processing and integration module and the sample transporter, if operable linkage to the sample transporter is desired can be used as known in the art or developed in the future. For some embodiments of the invention, particularly for plates with 96, 192, 384 and 864 wells per plate, detectors are available for integration into the system. Such detectors are described in U.S. Patent 5,589,351 (Harootunian),

20 U.S. Patent 5,355,215 (Schroeder), and PCT patent application WO 93/13423 (Akong). Alternatively, an entire plate may be "read" using an imager, such as a Molecular Dynamics Fluor-Imager 595 (Sunnyvale, CA). Multi-well platforms having greater than 864 wells, including 3,456 wells, can also be used in the present invention (see, for example, the PCT Application PCT/US98/11061, filed 6/2/98.

25 These higher density well plates require miniaturized assay volumes that necessitate the use of highly sensitivity assays that do not require washing. The present invention provides such assays as described herein.

The screening methods described herein can be made on cells growing in or deposited on solid surfaces. A common technique is to use a microtiter plate well

30 wherein the fluorescence measurements are made by commercially available fluorescent plate readers. One such method is to use cells in Costar 96 well microtiter

plates (flat with a clear bottom) and measure fluorescent signal with CytoFluor multiwell plate reader (Perseptive Biosystems, Inc., MA) using two emission wavelengths to record fluorescent emission ratios. In another embodiment, the system comprises a microvolume liquid handling system that uses electrokinetic forces to control the movement of fluids through channels of the system, for example as described in U.S. patent No., 5,800,690 issued September 1, 1998 to Chow *et al.*, European patent application EP 0 810 438 A2 filed May 5 1997, by Pelc *et al.* and PCT application WO 98/00231 filed 24 June 1997 by Parce *et al.* These systems use “chip” based analysis systems to provide massively parallel miniaturized analysis. Such systems are preferred systems of spectroscopic measurements in some instances that require miniaturized analysis.

*A method for identifying a chemical, modulator or a therapeutic*

The present invention can also be used for testing a therapeutic for useful therapeutic activity. A therapeutic is identified by contacting a test chemical suspected of having a modulating activity of a biological process or target with a test cell comprising the constructs of the present invention. Typically the cells are located within at least one well of a multi-well platform. The test chemical can be part of a library of test chemicals that is screened for activity, such as biological activity. The library can have individual members that are tested individually or in combination, or the library can be a combination of individual members. Such libraries can have at least two members, preferably greater than about 100 members or greater than about 1,000 members, more preferably greater than about 10,000 members, and most preferably greater than about 100,000 or 1,000,000 members. After appropriate incubation of the sample with the test cell, an inhibitor of protein synthesis may be added and a substrate for the reporter moiety added. At least one optical property (such as fluorescence or absorbance) of the sample is determined and compared to a non-treated control to determine the level of reporter gene expression or activity. If the sample having the test chemical exhibits increased or decreased reporter moiety expression or activity relative to that of the control cell then a candidate modulator has been identified.

The candidate modulator can be further characterized and monitored for structure, potency, toxicology, and pharmacology using well-known methods. The structure of a candidate modulator identified by the invention can be determined or confirmed by methods known in the art, such as mass spectroscopy. For putative modulators stored for extended periods of time, the structure, activity, and potency of the putative modulator can be confirmed.

Depending on the system used to identify a candidate modulator, the candidate modulator will have putative pharmacological activity. For example, if the candidate modulator is found to inhibit a protein tyrosine phosphatase involved, for example in T-cell proliferation *in vitro*, then the candidate modulator would have presumptive pharmacological properties as an immunosuppressant or anti-inflammatory (see, Suthanthiran *et al.*, (1996) Am. J. Kidney Disease, 28 159-172) Such nexuses are known in the art for several disease states, and more are expected to be discovered over time. Based on such nexuses, appropriate confirmatory *in vitro* and *in vivo* models of pharmacological activity, as well as toxicology, can be selected. The assays, and methods of use described herein, enable rapid pharmacological profiling to assess selectivity and specificity, and toxicity. This data can subsequently be used to develop new candidates with improved characteristics.

#### *Bioavailability and Toxicology of Candidate Modulators*

Once identified, candidate modulators can be evaluated for bioavailability and toxicological effects using known methods (see, Lu, Basic Toxicology, Fundamentals, Target Organs, and Risk Assessment, Hemisphere Publishing Corp., Washington (1985); U.S. Patent Nos: 5,196,313 to Culbreth (issued March 23, 1993) and U.S. Patent No. 5,567,952 to Benet (issued October 22, 1996). For example, toxicology of a candidate modulator can be established by determining *in vitro* toxicity towards a cell line, such as a mammalian i.e. human, cell line. Candidate modulators can be treated with, for example, tissue extracts, such as preparations of liver, such as microsomal preparations, to determine increased or decreased toxicological properties of the chemical after being metabolized by a whole organism. The results of these

types of studies are often predictive of toxicological properties of chemicals in animals, such as mammals, including humans.

The toxicological activity can be measured using reporter genes that are activated during toxicological activity or by cell lysis (see WO 98/13353, published 4/2/98). Preferred reporter genes produce a fluorescent or luminescent translational product (such as, for example, a Green Fluorescent Protein (see, for example, U.S. Patent No. 5,625,048 to Tsien *et al.*, issued 4/29/98; U.S. Patent No. 5,777,079 to Tsien *et al.*, issued 7/7/98; WO 96/23810 to Tsien, published 8/8/96; WO 97/28261, published 8/7/97; PCT/US97/12410, filed 7/16/97; PCT/US97/14595, filed 8/15/97)) or a translational product that can produce a fluorescent or luminescent product (such as, for example, beta-lactamase (see, for example, U.S. Patent No. 5,741,657 to Tsien, issued 4/21/98, and WO 96/30540, published 10/3/96)), such as an enzymatic degradation product. Cell lysis can be detected in the present invention as a reduction in a fluorescence signal from at least one photon-producing agent within a cell in the presence of at least one photon reducing agent. Such toxicological determinations can be made using prokaryotic or eukaryotic cells, optionally using toxicological profiling, such as described in PCT/US94/00583, filed 1/21/94 (WO 94/17208), German Patent No 69406772.5-08, issued 11/25/97; EPC 0680517, issued 11/12/94; U.S. Patent No. 5,589,337, issued 12/31/96; EPO 651825, issued 1/14/98; and U.S. Patent No. 5,585,232, issued 12/17/96).

Alternatively, or in addition to these *in vitro* studies, the bioavailability and toxicological properties of a candidate modulator in an animal model, such as mice, rats, rabbits, or monkeys, can be determined using established methods (see, Lu, supra (1985); and Creasey, Drug Disposition in Humans, The Basis of Clinical Pharmacology, Oxford University Press, Oxford (1979), Osweiler, Toxicology, Williams and Wilkins, Baltimore, MD (1995), Yang, Toxicology of Chemical Mixtures; Case Studies, Mechanisms, and Novel Approaches, Academic Press, Inc., San Diego, CA (1994), Burrell *et al.*, Toxicology of the Immune System; A Human Approach, Van Nostrand Reinhold, Co. (1997), Niesink *et al.*, Toxicology; Principles and Applications, CRC Press, Boca Raton, FL (1996)). Depending on the toxicity, target organ, tissue, locus, and presumptive mechanism of the candidate modulator,

the skilled artisan would not be burdened to determine appropriate doses, LD<sub>50</sub> values, routes of administration, and regimes that would be appropriate to determine the toxicological properties of the candidate modulator. In addition to animal models, human clinical trials can be performed following established procedures, such as those set forth by the United States Food and Drug Administration (USFDA) or equivalents of other governments. These toxicity studies provide the basis for determining the therapeutic utility of a candidate modulator *in vivo*.

### *Efficacy of Candidate Modulators*

Efficacy of a candidate modulator can be established using several art-recognized methods, such as *in vitro* methods, animal models, or human clinical trials (see, Creasey, supra (1979)). Recognized *in vitro* models exist for several diseases or conditions. For example, the ability of a chemical to extend the life-span of HIV-infected cells *in vitro* is recognized as an acceptable model to identify chemicals expected to be efficacious to treat HIV infection or AIDS (see, Daluge *et al.*, (1995) Antimicro. Agents Chemother. 41 1082-1093). Furthermore, the ability of cyclosporin A (CsA) to prevent proliferation of T-cells *in vitro* has been established as an acceptable model to identify chemicals expected to be efficacious as immunosuppressants (see, Suthanthiran *et al.*, supra, (1996)). For nearly every class of therapeutic, disease, or condition, an acceptable *in vitro* or animal model is available. Such models exist, for example, for gastro-intestinal disorders, cancers, cardiology, neurobiology, and immunology. In addition, these *in vitro* methods can use tissue extracts, such as preparations of liver, such as microsomal preparations, to provide a reliable indication of the effects of metabolism on the candidate modulator. Similarly, acceptable animal models may be used to establish efficacy of chemicals to treat various diseases or conditions. For example, the rabbit knee is an accepted model for testing chemicals for efficacy in treating arthritis (see, Shaw and Lacy, J. (1973) Bone Joint Surg. (Br) 55 197-205. Hydrocortisone, which is approved for use in humans to treat arthritis, is efficacious in this model which confirms the validity of this model (see, McDonough, (1982) Phys. Ther. 62 835-839). When choosing an appropriate model to determine efficacy of a candidate modulator, the skilled artisan

can be guided by the state of the art to choose an appropriate model, dose, and route of administration, regime, and endpoint and as such would not be unduly burdened.

In addition to animal models, human clinical trials can be used to determine the efficacy of a candidate modulator in humans. The USFDA, or equivalent governmental agencies, have established procedures for such studies (see, [www.fda.gov](http://www.fda.gov)).

#### *Selectivity of Candidate Modulators*

The *in vitro* and *in vivo* methods described above also establish the selectivity of a candidate modulator. It is recognized that chemicals can modulate a wide variety of biological processes or be selective. Panels of cells, each containing constructs with varying specificity, based on the present invention, can be used to determine the specificity of the candidate modulator. Selective modulators are preferable because they have fewer side effects in the clinical setting. The selectivity of a candidate modulator can be established *in vitro* by testing the toxicity and effect of a candidate modulator on a plurality of cell lines that exhibit a variety of cellular pathways and sensitivities. The data obtained from these *in vitro* toxicity studies can be extended into *in vivo* animal model studies, including human clinical trials, to determine toxicity, efficacy, and selectivity of the candidate modulator using art-recognized methods.

#### *An identified chemical, modulator, or therapeutic and compositions*

The invention includes compositions, such as novel chemicals, and therapeutics identified by at least one method of the present invention as having activity by the operation of methods, systems or components described herein. Novel chemicals, as used herein, do not include chemicals already publicly known in the art as of the filing date of this application. Typically, a chemical would be identified as having activity from using the invention and then its structure revealed from a proprietary database of chemical structures or determined using analytical techniques such as mass spectroscopy.

One embodiment of the invention is a chemical with useful activity, comprising a chemical identified by the method described above. Such compositions include small organic molecules, nucleic acids, peptides and other molecules readily synthesized by techniques available in the art and developed in the future. For example, the following combinatorial compounds are suitable for screening: peptoids (PCT Publication No. WO 91/19735, 26 Dec. 1991), encoded peptides (PCT Publication No. WO 93/20242, 14 Oct. 1993), random bio-oligomers (PCT Publication WO 92/00091, 9 Jan. 1992), benzodiazepines (U.S. Patent No. 5,288,514), diversomeres such as hydantoins, benzodiazepines and dipeptides (Hobbs DeWitt, S. *et al.*, (1993) Proc. Nat. Acad. Sci. USA 90 6909-6913), vinylogous polypeptides (Hagihara *et al.*, (1992) J. Amer. Chem. Soc. 114 6568), nonpeptidal peptidomimetics with a Beta-D-Glucose scaffolding (Hirschmann, R. *et al.*, (1992) J. Amer. Chem. Soc. 114 9217-9218), analogous organic syntheses of small compound libraries (Chen, C. *et al.*, (1994) J. Amer. Chem. Soc. 116 2661), oligocarbamates (Cho, C.Y. *et al.*, (1993) Science 261: 1303), and/or peptidyl phosphonates (Campbell, D.A. *et al.*, (1994) J. Org. Chem. 59 658). See, generally, Gordon, E. M. *et al.*, (1994). J. Med Chem. 37 1385. The contents of all of the aforementioned publications are incorporated herein by reference.

The present invention also encompasses the identified compositions in a pharmaceutical composition comprising a pharmaceutically acceptable carrier prepared for storage and subsequent administration, which have a pharmaceutically effective amount of the products disclosed above in a pharmaceutically acceptable carrier or diluent. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro edit. 1985). Preservatives, stabilizers, dyes and even flavoring agents may be provided in the pharmaceutical composition. For example, sodium benzoate, ascorbic acid and esters of p-hydroxybenzoic acid may be added as preservatives. In addition, antioxidants and suspending agents may be used.

The compositions of the present invention may be formulated and used as tablets, capsules or elixirs for oral administration; suppositories for rectal

administration; sterile solutions, suspensions for injectable administration; and the like. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Suitable excipients are, for example, water, saline, dextrose, mannitol, lactose, lecithin, albumin, sodium glutamate, cysteine hydrochloride, and the like. In addition, if desired, the injectable pharmaceutical compositions may contain minor amounts of nontoxic auxiliary substances, such as wetting agents, pH buffering agents, and the like. If desired, absorption enhancing preparations (e.g., liposomes) may be utilized.

The pharmaceutically effective amount of the composition required as a dose will depend on the route of administration, the type of animal being treated, and the physical characteristics of the specific animal under consideration. The dose can be tailored to achieve a desired effect, but will depend on such factors as weight, diet, concurrent medication and other factors which those skilled in the medical arts will recognize. In practicing the methods of the invention, the products or compositions can be used alone or in combination with one another or in combination with other therapeutic or diagnostic agents. These products can be utilized *in vivo*, ordinarily in a mammal, preferably in a human, or *in vitro*. In employing them *in vivo*, the products or compositions can be administered to the mammal in a variety of ways, including parenterally, intravenously, subcutaneously, intramuscularly, colonically, rectally, nasally or intraperitoneally, employing a variety of dosage forms. Such methods may also be applied to testing chemical activity *in vivo*.

As will be readily apparent to one skilled in the art, the useful *in vivo* dosage to be administered and the particular mode of administration will vary depending upon the age, weight and mammalian species treated, the particular compounds employed, and the specific use for which these compounds are employed. The determination of effective dosage levels, that is the dosage levels necessary to achieve the desired result, can be accomplished by one skilled in the art using routine pharmacological methods. Typically, human clinical applications of products are commenced at lower dosage levels, with dosage level being increased until the desired effect is achieved. Alternatively, acceptable *in vitro* studies can be used to



establish useful doses and routes of administration of the compositions identified by the present methods using established pharmacological methods.

In non-human animal studies, applications of potential products are commenced at higher dosage levels, with dosage being decreased until the desired effect is no longer achieved or adverse side effects disappear. The dosage for the products of the present invention can range broadly depending upon the desired affects and the therapeutic indication. Typically, dosages may be between about 10 mg/kg and 100 mg/kg body weight, and preferably between about 100 µg/kg and 10 mg/kg body weight. Administration is preferably oral on a daily basis.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl *et al.*, in *The Pharmacological Basis of Therapeutics*, 1975). It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in *Remington's Pharmaceutical Sciences*, 18th Ed., Mack Publishing Co., Easton, PA (1990). Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For such transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art. Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external micro-environment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for

oral administration may be in the form of tablets, dragees, capsules, or solutions. The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, for example, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, 5 entrapping, or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or 10 synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated 15 solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as 20 sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, 25 agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to 30 the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. For this purpose, concentrated sugar

solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. Such formulations can be made using methods known in the art (see, for example, U.S. Patent Nos. 5,733,888 (injectable compositions); 5,726,181 (poorly water soluble compounds); 5,707,641 (therapeutically active proteins or peptides); 5,667,809 (lipophilic agents); 5,576,012 (solubilizing polymeric agents); 5,707,615 (anti-viral formulations); 5,683,676 (particulate medicaments); 5,654,286 (topical formulations); 5,688,529 (oral suspensions); 5,445,829 (extended release formulations); 5,653,987 (liquid formulations); 5,641,515 (controlled release formulations) and 5,601,845 (spheroid formulations).

## VII. TRANSGENIC ANIMALS

In another embodiment, the invention provides a transgenic non-human organism that expresses a nucleic acid sequence that encodes a target protein, (such as a reporter moiety, enzyme or structural protein) functionally coupled to one or more destabilization domains by a linker. Because such constructs can be expressed within intact living cells, with preset degrees of stability, the invention provides the ability to regulate the expression level of the target protein, or to monitor post translational activities within defined cell populations, tissues or in an entire transgenic organism.

In one embodiment the approach may be used to regulate the expression level of an enzyme or group of enzymes involved in a particular signal transduction, disease, or metabolic pathway. Such methods may be useful, for example, for creating transgenic model animals for certain disease states, or for modulating the intracellular concentration of enzymatic intermediates through the manipulation of the expression levels of the enzymes involved. For example, to increase the intracellular concentration of an intermediate one could increase the concentration of the enzyme(s) involved in the synthesis of the intermediate, and / or decrease the concentration of the enzyme(s) involved in degradation of the intermediate. Typically

the approach would require the replacement of the native enzymes with fusion proteins of the enzymes with the multimerized destabilization domains of the present invention. For target proteins in which the desired concentration was relatively high, one would select fusion proteins with relatively few (i.e. one or two), or even no, (zero) copies of the destabilization domain. For target proteins for which a relatively low intracellular concentration was desired, one would select fusion proteins with relatively more copies of the destabilization domain (i.e. three or more).

In another embodiment, the approach can be used to identify where in specific tissues a particular activity is located, for example, by expression of a reporter moiety coupled to the multimerized destabilization domain via a linker comprising recognition and cleavage motifs for that activity, in the organism. Typically the linker would comprise a single polypeptide chain that covalently couples the destabilization domains to the reporter moiety. Typically in this embodiment, the linker will comprise a post-translational recognition motif such as a protease recognition motif. Cleavage of the linker by the protease at the cleavage site results in uncoupling of the multimerized destabilization domains from the reporter moiety resulting in a modulation in the stability of the reporter moiety, thereby resulting in an accumulation of reporter moiety in cells or tissues that exhibit protease activity.

Such non-human organisms include vertebrates such as rodents, fish such as Zebrafish, non-human primates and reptiles as well as invertebrates. Preferred non-human organisms are selected from the rodent family including rat and mouse, most preferably mouse. The transgenic non-human organisms of the invention are produced by introducing transgenes into the germline of the non-human organism. Embryonic target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the organism and stage of development of the embryonic target cell. In vertebrates, the zygote is the best target for microinjection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter, which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82

4438-4442.). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

A transgenic organism can be produced by cross-breeding two chimeric organisms which include exogenous genetic material within cells used in reproduction. Twenty-five percent of the resulting offspring will be transgenic *i.e.*, organisms that include the exogenous genetic material within all of their cells in both alleles. 50% of the resulting organisms will include the exogenous genetic material within one allele and 25% will include no exogenous genetic material.

Retroviral infection can also be used to introduce transgene into a non-human organism. In vertebrates, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retro viral infection (Jaenich, R., (1976) Proc. Natl. Acad. Sci USA 73 1260-1264.). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan, *et al.* (1986) in Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The viral vector system used to introduce the transgene is typically a replication-defective retro virus carrying the transgene (Jahner, *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82 6927-6931; Van der Putten, *et al.*, (1985) Proc. Natl. Acad. Sci USA 82 6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart, *et al.*, (1987) EMBO J. 6 383-388).

Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (D. Jahner *et al.*, (1982) Nature 298 623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells that formed the transgenic nonhuman animal. Further, the founder may contain various retro viral insertions of the transgene at different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit with low

efficiency, by intrauterine retro viral infection of the midgestation embryo (D. Jahner *et al.*, *supra*). A third type of target cell for transgene introduction for vertebrates is the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (M. J. Evans *et al.* (1981) *Nature* 292 154-156; M.O. Bradley *et al.*, (1984) *Nature* 309 255-258; Gossler, *et al.*, (1986) *Proc. Natl. Acad. Sci USA* 83 9065-9069; and Robertson *et al.*, (1986) *Nature* 322 445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retro virus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. (For review see Jaenisch, R., (1988) *Science* 240 1468-1474).

## VIII TRANSGENIC PLANTS

In another embodiment, the invention provides a transgenic plant that expresses a nucleic acid sequence that encodes a target protein, (such as a reporter moiety, enzyme or structural protein) functionally coupled to a multimerized destabilization domain by a linker. Because such constructs can be specifically expressed, both spatially and temporally, within intact living cells, the invention provides the ability to regulate the expression level of the target protein, within defined cell populations, tissues, or in the entire transgenic plant.

In one embodiment the approach may be used to regulate the expression level of an enzyme or group of enzymes involved in a particular signal transduction, developmental or metabolic pathway. Such methods may be useful for creating transgenic plants with improved disease resistance or other favorable traits. More particularly, plants can be genetically engineered to express various phenotypes of agronomic interest, for example by allowing for the regulated expression of agronomically important genes. Given potential concerns about the safety of transgenic plants, the ability to reduce or eliminate the expression of certain resistance genes prior to harvesting and human consumption is of particular interest. Examples of the types of genes that could be manipulated using the methods described herein, include disease resistance genes, herbicide resistance genes and genes that improve plant traits, including those shown in **Table 4**, below.

<b>TABLE 4</b>		
<b>I. Disease Resistance Genes</b>		
<b>Gene or Gene Product</b>	<b>Function</b>	<b>Reference</b>
Tomato Cf-9 gene	Resistance to <i>Cladosporium fulvum</i>	Jones et al., Science 266 789 (1994)
Tomato Pto gene	Resistance to <i>Pseudomonas syringae</i>	Martin et al., Science 262: 1432 (1993)
Arabidopsis RSP2 gene	Resistance to <i>Pseudomonas syringae</i>	Mindrin et al., Cell 78: 1089 (1994)
Bacillus thuringiensis protein	Insect resistance	Geiser et al., Gene 48: 109 (1986),
<i>Streptomyces nitrosporeus</i> $\alpha$ -amylase inhibitor	Inhibition of amylase activity.	Sumitani et al., Biosci. Biotech. Biochem. 57 1243 (1993)
Expression of insect-specific	Disruption of insect	Hammock et al., Nature 344:



hormones or pheromones such as an ecdysteroid and juvenile hormone	development	458 (1990)
Expression insect-specific scorpion venom	Insect resistance	Pang et al., Gene 116: 165 (1992)
Altered expression of metabolic enzymes	Expression of enzymes responsible for the formation of non protein molecules with insecticidal activity	
Altered expression of signal transduction enzymes	Expression of enzymes responsible for the post-translational modification of biologically active molecules	See PCT application WO 93/02197, Botella et al., Plant Molec. Biol. 24: 757 (1994),
Expression of synthetic antimicrobial peptides, such as peptide derivatives of Tachyplesin	Improved disease resistance	
Altered expression of Ion channels, blockers or permeases such as cecropin-3 lytic peptide	Improved resistance to Pseudomonas solanacearum.	Jaynes et al., Plant Sci. 89: 43 (1993),
Expression of viral coat proteins or viral-invasive proteins or toxins.	Improved viral resistance to alfalfa mosaic virus, cucumber mosaic virus, tobacco streak virus, potato virus X, potato virus Y, tobacco etch virus, tobacco rattle virus and tobacco mosaic virus	See Beachy et al., Ann. Rev. Phytopathol. 28: 451 (1990).
Expression of insect-specific antibody or immunotoxins	Improved resistance to insects	Taylor et al., Abstract #497, SEVENTH INT'L SYMPOSIUM ON MOLECULAR PLANT-MICROBE INTERACTIONS (1994)
Expression of virus-specific antibodies.	Improved resistance to viruses	Tavladoraki et al., Nature 366: 469 (1993)
Expression of developmental - arrestive proteins or gene products, as endo α1,4-D-polygalacturonase, or expression of barley ribosome-inactivating gene	Increased resistance to pathogens or parasites	See Lamb et al., Biol Technology ;Q: 1436 (1992).  Logemann et al., BiolTechnology.10: 30 (1992)
<b>II. Herbicide Resistance Genes</b>		
Expression of mutant ALS and AHAS enzymes	Inhibition of the growing point or meristem, increasing resistance to herbicides	Lee et al., EMBO J. 7: 1241 (1988), and Miki et al., Theor. Appl. Genet. 8 : 449 (1990),
Expression of mutant EPSP synthase and aroA genes,	Resistance to glyphosate and other phosphono compounds such as glufosinate	U.S. patent No. 4,940,835 to Shah et al., U.S. patent No. 4,769,061 to Comai. European patent application No. 0 333 033 to Kumada et al. and U.S. patent No. 4,975,374 to Goodman et al.
<b>III. Genes That Confer Or Contribute To A Value-added Trait</b>		
Expression of antisense gene of	Improved fatty acid composition	Knultzon et al., Proc. Natl.

stearoyl-ACP desaturase		Acad. Sci.USA 89: 2624 (1992).
Expression of phytic acid degrading enzymes	Improved free phosphate composition	Van Hartingsveldt et al., Gene 127: 87 (1993)
Expression of fructosyltransferase, levansucrase, or invertase genes	Improved carbohydrate composition	See Shiroza et al., J. Bacteriol. 170: 810 (1988), Steinmetz et al., Mol. Gen. Genet. 200 220 (1985), Elliot et al., Plant Molec. Biol. 21 515 (1993)

In another embodiment, the approach can be used to specifically identify where in specific tissues a particular activity is expressed, for example by expression of the protease sensor in specific plant tissues.

5 Transgenic plants may be produced by any one of a number of methods of plant transformation and regeneration. Numerous methods for plant transformation have been developed, including biological and physical, plant transformation protocols. See, for example, Miki et al., "Procedures for Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc. , Boca Raton, 1993) pages 67-88. In  
10 addition, expression vectors and in vitro culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber et al., "Vectors for Plant Transformation" in Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pages 89-119.

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of Agrobacterium. See, for example, Horsch et al., (1985) Science 227 1229. A. tumefaciens and A. rhizogenes are plant pathogenic soil bacteria which genetically transform plant cells. The Ti  
20 and Ri plasmids of A. tumefaciens and A. rhizogenes, respectively, carry genes responsible for genetic transformation of the plant See, for example, Kado, C.I., Crit. Rev. Plant. Sci. 10: 1 (1991). Descriptions of Agrobacterium vector systems and methods for Agrobacterium-mediated gene transfer are provided by Gruber et al., supra, Miki et al., supra, and Moloney et al., (1989) Plant Cell Reports 8 238.

25 Despite the fact the host range for Agrobacterium mediated transformation is broad, some major cereal crop species and gymnosperms have generally been

recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice. Hiei et al., (1994) *The Plant Journal* 6 271-282. Several methods of plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to *Agrobacterium*-mediated transformation.

5        A generally applicable method of plant transformation is microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles measuring 1 to 4  $\mu$ m. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate plant cell walls and membranes. Sanford et al., (1987),  
10    Part. Sci. Technol. 5 27, Sanford, J.C., (1988) *Trends Biotech.* 6 299, Sanford, J.C., (1990) *Physiol. Plant* 79 206, Klein et al., (1992) *Biotechnology* 10 268.

Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., (1991) *BioTechnology* 9 996. Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants.  
15    Deshayes et al., (1985) *EMBO J.*, 4 2731, Christou et al., (1987) *Proc Natl. Acad. Sci. U.S.A.* 84 3962. Direct uptake- of DNA into protoplasts using  $\text{CaCl}_2$  precipitation, polyvinyl alcohol or poly-L-ornithine have also been reported. Hain et al., (1985) *Mol.Gen. Genet.* 199 161 and Draper et al., (1982) *Plant Cell Physiol.* 23 451. Electroporation of protoplasts and whole cells and tissues have also been  
20    described. Donn et al., In Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, p 53 (1990) ; D'Halluin et al., (1992) *Plant Cell* 4 1495-1505 and Spencer et al., (1994) *Plant Mol. Biol.* 24 51-61.

A preferred method is microprojectile-mediated bombardment of immature embryos. The embryos can be bombarded on the embryo axis side to target the  
25    meristem at a very early stage of development or bombarded on the scutellar side to target cells that typically form callus and somatic embryos. Targeting of the scutellum using projectile bombardment is well known to those in the art of cereal tissue culture. Klein et al., (1988) *BioTechnol.* , 6 559-563; Sautter et al., *BiolTechnol.*, 9 1080-1085 (1991) ; Chibbar et al., (1991) *Genome*, 34 435-460. The scutellar origin of  
30    regenerable callus from cereals is well known. Green et al., (1975) *Crop Sci.*, 15 417-421; Lu et al., (1982) *TAG* 62 109-112; and Thomas and Scott, (1985) *J. Plant Physiol.* 121 159-169 - Targeting the scutellum and then using chemical selection to

recover transgenic plants is well established in cereals. D/Halluin et al., Plant Cell 4: 1495-1505 (1992) ; Perl et al., MGG 235: 279-284 (1992); Cristou et al., BiolTechnol. 9: 957-962 (1991). This literature reports DNA targeting of the scutellum and recovery of transgenic callus, plants and progeny based on a chemical selection regime. None of these references teach successful plant transformation wherein transformed cells are visualized with a screenable marker such as GUS.

A preferred transformation method involves bombardment of the scutellar surface of immature embryos to introduce the expression cassette with the gene for a bioluminescent protein, such as *Aequorea victoria* GFP (See PCT publication WO 97/41228 to Gordon-Kamm et al., incorporated herein by reference). Embryos can be pretreated for 1 to 48 hours with a high osmoticum medium or left on a highosmoticum medium for 24-48 hours after bombardment to improve cell survival and transformation frequencies. Immature embryos are then cultured on typical callusinducing medium with no selective agent. At each subculture transfer, i.e., every two weeks, the culture is monitored using UV-blue light for GFP fluorescence. Fluorescing calli are separated from non-fluorescing callus, and grown to the point where plants can be regenerated through standard media progressions.

Plants can be manipulated, for example, by removal of the apical meristem, to stimulate axillary or secondary buds which can exhibit larger transgenic sectors relative to the primary shoot. Flowers above transgenic shoots are pollinated and the progeny are analyzed for transgene presence and expression. A variety of starting explants can regenerate shoots in sunflower, and thus represent alternative targets for GFP-encoding DNA delivery and transmission to progeny. These include the seedling meristem (as above), also the seedling hypocotyl, the mature cotyledon, the immature cotyledon, zygotic immature embryos, somaticembryos, and primary leaflets. See for example, respectively, Greco et al., (1984) Plant Sci. Lett. 36 73-77; Krauter et al., (1991) Helia 14 117-122; Power (1987) Am. J. Bot. 74 497503; Krauter et al., (1991) Theor. Appl. Genet. a2: 521525; Finer, (1987) Plant Cell Rep. J: 372-374, and Greco et al., (1984) Plant Sci. Lett. 36 73-77.

30

## EXAMPLES

### Example 1 Generation of multimerized destabilization domains

The cDNA encoding human ubiquitin was isolated from a human genomic DNA preparation obtained from Jurkat cells by polymerase chain reaction (PCR) using the PCR primers Ubi5 (SEQ. ID. NO. 15) and Ubi3 (SEQ. ID. NO. 16) and cloned into pBluescript II vector (Stratagene). The C-terminal residue of ubiquitin was altered from glycine to valine by site-directed mutagenesis (Kunkel) in order to generate a mutant form of ubiquitin that cannot be cleaved by cellular  $\alpha$ -NH-ubiquitin endopeptidases. This mutant is hereafter referred to as ubiquitinG76V (SEQ. ID. NO. 17). The ubiquitinG76V (SEQ. ID. NO. 17) mutant was then amplified by PCR using the oligonucleotide primers Ub5' (SEQ. ID. NO. 18 ) and Ub3', (SEQ. ID. NO. 19). These primers introduce a Bgl II restriction site at the 5' end of the coding sequence and a BamH I site at the 3' end of the coding sequence. The PCR fragment from the reaction was digested with Bgl II and BamH I and ligated into BamH I-digested pBluescript II vector. This plasmid was then digested with Bgl II and BamH I and the ubiquitinG76V (SEQ. ID. NO. 17) containing fragment was isolated and ligated to generate multimerized ubiquitinG76V domains. The ubiquitinG76V multimers were digested with Bgl II and BamH I to ensure that the individual ubiquitinG76V domains (SEQ. ID. NO. 17) were in the correct orientation. The digested ubiquitinG76V multimers were separated by agarose gel electrophoresis and multimers of the appropriate sizes were isolated and cloned into BamH I-digested pBluescript II. The ubiquitinG76V multimers were then excised using BamH I and Hind III and subcloned to generate a series of plasmids containing in frame fusions of from one to four copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to the reporter moiety or protein of interest. These constructs are referred to as 1XUb (one copy of ubiquitinG76V (SEQ. ID. NO. 17)), 2XUb (two copies of ubiquitinG76V (SEQ. ID. NO. 17)), 3XUb (three copies of ubiquitinG76V (SEQ. ID. NO. 17)) and 4XUb (four copies of ubiquitinG76V (SEQ. ID. NO. 17)).

30

## Example 2. Creation of multimerized destabilization domain-β-lactamase fusion proteins

The gene encoding the E. coli TEM-1 β-lactamase was isolated from the plasmid pBluescript (Stratagene) by polymerase chain reaction (PCR) amplification using the PCR primers BLA5 (SEQ. ID. NO. 20 ) and ABSC107, (SEQ. ID. NO. 21) resulting in the deletion of the signal sequence and introduction of a BamH I restriction site and the amino acids below at the 5' end of the coding sequence.

BamH1  
H G S G A W L **H P E T L V K V K**

Amino acids in bold represent original β-lactamase coding sequence, underlined amino acids represent the BamH I restriction site. An Xba I site was inserted at the 3' end of the coding sequence. The PCR fragments from these reactions were digested with BamH I and Xba I and ligated into pcDNA3 (Invitrogen) via the same sites. The resulting construct, pcDNA3-Bla (SEQ. ID. NO. 22), was then used to create in-frame fusions with the multimerized ubiquitinG76V constructs above. This was achieved by digesting the multimerized ubiquitinG76V constructs with the restriction enzymes BamH I and Hind III, and then ligating them via the same sites into the pcDNA3-Bla construct. These constructs were named pcDNA3-1XUb-Bla (SEQ. ID. NO. 23), pcDNA3-2XUb-Bla (SEQ. ID. NO. 24), pcDNA3-3XUb-Bla (SEQ. ID. NO. 25), pcDNA3-4XUb-Bla (SEQ. ID. NO. 26). To produce the wild-type β-lactamase protein, we used a construct that contains one copy of wild-type (cleavable) ubiquitin (SEQ. ID. NO. 2) fused to the β-lactamase coding region in the pcDNA3 vector; this plasmid is referred to as pcDNA3-Ub-Met-Bla (SEQ. ID. NO. 27). Upon synthesis of the Ub-Met-Bla fusion protein, ubiquitin isopeptidases efficiently cleave off the N-terminal ubiquitin (SEQ. ID. NO. 2) precisely after glycine-76, generating the wild-type β-lactamase protein with methionine at its N-terminus.

## Example 3. Creation of multimerized destabilization domain-Naturally Fluorescent Protein fusions

The gene encoding the GFP mutant Emerald (S65T, S72A, N149K, M153T, I167T) (SEQ. ID. NO. 28) was amplified by PCR using the oligonucleotides GFP5' (SEQ. ID. NO. 29) and GFP3', (SEQ. ID. NO. 30). The resulting PCR product had a BamH I restriction site at the 5' end of the coding sequence and a Xba I site at the 3' end of the coding sequence. The PCR fragment from this reaction was digested with BamH I and Xba I and ligated into pcDNA3 via the same sites. The resulting construct, pcDNA3-GFP was then used to create in-frame fusions with the multimerized ubiquitinG76V constructs described above. This was achieved by digesting the pcDNA3-1-4XUb-Bla constructs (SEQ. ID. NOs. 23 to 26) with the restriction enzymes BamH I and Hind III, and then ligating the fragment encoding the various multiUb destabilization domains via the same sites into the pcDNA3-GFP construct. These constructs were named pcDNA3-1XUb-GFP (SEQ. ID. NO. 31), pcDNA3-2XUb-GFP (SEQ. ID. NO. 32), pcDNA3-3XUb-GFP (SEQ. ID. NO. 33), pcDNA3-4XUb-GFP (SEQ. ID. NO. 34).

#### **Example 4. Creation of multimerized destabilization domain-Naturally Occurring Mammalian Protein fusions**

Fusions between multimerized uncleavable ubiquitinG76V (SEQ. ID. NO. 17) and caspase-3 were constructed to further investigate the relationship between the degree of destabilization exerted by varying the number of copies of the destabilization domain with different target proteins.

The caspase-3 cDNA (SEQ. ID. NO. 35) was amplified by PCR using the primers C35' (SEQ. ID. NO. 36) and C33' (SEQ. ID. NO. 37) to add BamH I sites at the ends of the caspase-3 cDNA. The amplified caspase-3 cDNA was digested with BamH I then cloned into BamH I-digested pcDNA3-1-4XUb-Bla plasmids (SEQ. ID. NOs. 23 to 26), to create fusions of the different multiubiquitin destabilization domains to a caspase-3- $\beta$ -lactamase fusion. The  $\beta$ -lactamase coding region was then removed from these plasmids by digesting to completion with Xba I followed by a partial digest with BamH I. The digests were separated by agarose gel electrophoresis and the correct size DNA band was purified from the gel. The ends of the digested plasmid were blunted with the Klenow fragment of DNA polymerase and the plasmid recircularized by ligation. The resulting plasmids contained an in-frame fusion of the

ubiquitinG76V destabilization domain (with from one to four copies of ubiquitinG76V (SEQ. ID. NO. 17)) to the caspase-3 coding region. These plasmids were designated pcDNA3-1-4XUb-C3 (SEQ. ID. NO. 38 to 41). To produce the wild-type caspase-3 protein, the caspase-3 cDNA was amplified by PCR with primers C35Met (SEQ. ID. NO. 42) and C33' (SEQ. ID. NO. 43) and cloned directly into pcDNA3-Ub-Met-Bla (SEQ. ID. NO. 27). The resulting plasmid was then digested with BamH I and Xba I and recircularized as described above to create the wild-type caspase-3 control construct; this plasmid was designated as pcDNA3-Ub-Met-C3 (SEQ. ID. NO. 44). Upon synthesis of the Ub-Met-caspase-3 fusion protein, ubiquitin isopeptidases efficiently cleave off the N-terminal ubiquitin precisely after glycine-76, generating the wild-type caspase-3 protein with methionine at its N-terminus (data not shown).

**Example 5. Characterization of multimerized destabilization domain- $\beta$ -lactamase fusion proteins *in vitro*.**

<sup>35</sup>S-Labeled multimerized destabilization domain- $\beta$ -lactamase fusion protein molecules were produced using a coupled *in vitro* transcription/translation system based on a rabbit reticulocyte lysate (TNT T7 Quick; Promega). Constructs containing from one to four copies of the destabilization domain (pcDNA3-1-4XUb-Bla (SEQ. ID. NOs. 23 to 26) from Example 2) were incubated in the TNT lysate essentially as described in the manufacturer's directions in the presence of 0.25 mCi/ml <sup>35</sup>S-methionine (10 mCi/ml, 1175 Ci/mmol; New England Nuclear) to generate <sup>35</sup>S-labeled fusion proteins.

To determine the half life of the constructs, 1  $\mu$ l samples of the synthesis reactions were incubated at 37°C in 9  $\mu$ l of chase extract (crude rabbit reticulocyte lysate (Promega) supplemented with 100  $\mu$ g/ml cycloheximide, 1 mM ATP, 20 mM phosphocreatine, 2.5 mM MgCl<sub>2</sub>, 5  $\mu$ g/ml creatine kinase, 200  $\mu$ g/ml ubiquitin, and 50  $\mu$ M methionine). The rabbit reticulocyte lysate system contains all of the components necessary for efficient recognition and degradation of proteins by the ubiquitin-proteasome pathway. Samples were removed at 0, 5, 10, 20, 30, 45 and 60 minutes of reaction and analyzed by polyacrylamide gel electrophoresis (SDS-



PAGE). The gels were treated with Amplify (Amersham) and the labeled species detected by autoradiography. This analysis showed that wild-type  $\beta$ -lactamase was stable over the 1 hour chase period while the ubiquitinG76V- $\beta$ -lactamase fusions were considerably less stable (**FIG. 2A**). In particular, the 1XUb-Bla fusions were modestly destabilized ( $t_{1/2} \sim 20$  min) and  $\beta$ -lactamase fusions containing 2, 3 or 4 copies of ubiquitinG76V (SEQ. ID. NO. 17) were strongly destabilized ( $t_{1/2} < 5$  min). In addition, the degradation of the 2XUb-Bla fusion was slightly slower than the degradation of  $\beta$ -lactamase fusions containing 3 or 4 copies of ubiquitinG76V (SEQ. ID. NO. 17) (**FIG. 2A**).

In order to test whether the degradation of multiUb-Bla fusions *in vitro* is dependent on the proteasome, TNT synthesis reactions were performed in the absence or presence of the proteasome inhibitor MG132 (Calbiochem) at 50  $\mu$ M and analyzed by SDS-PAGE as described above. These experiments showed that inhibition of the proteasome resulted in a dramatic increase in the amount of fusion protein synthesized for  $\beta$ -lactamase fusions containing 2, 3 or 4 copies of ubiquitinG76V (SEQ. ID. NO. 17) while MG132 had very little or no significant effect on the synthesis of wild-type  $\beta$ -lactamase or 1XUb-Bla (**FIG. 2B**). Use of MG132 in these *in vitro* reactions also revealed the presence of labeled high molecular weight species that represent extended ubiquitin chains conjugated to the ubiquitinG76V- $\beta$ -lactamase fusions (also see Example 16). Therefore, the uncleavable ubiquitinG76V domains (SEQ. ID. NO. 17) in the multiubiquitin destabilization domain may be acting as high affinity conjugation sites for further ubiquitination by E2/E3 ubiquitin ligases. The relative lack of these high molecular weight species in the absence of MG132 reflects the highly efficient recognition and degradation by the proteasome of proteins tagged with extended polyubiquitin chains.

#### **Example 6. Characterization of multimerized destabilization domain-Naturally Fluorescent Protein fusions *in vitro*.**

Characterization of the turnover of multiubiquitin-GFP fusion proteins *in vitro* was similar to the multiubiquitin- $\beta$ -lactamase analyses described in Example 5, except that time points were taken at 0, 30, 60, 90 and 120 min. These experiments showed that Emerald GFP (SEQ. ID. NO. 28) is extremely stable under these conditions, and that the multiubiquitin destabilization domain was able to impart a short half-life upon the multiUb-GFP fusion proteins (**FIG. 3**). A striking feature of this analysis was that significant destabilization of GFP required higher numbers of ubiquitinG76V (SEQ. ID. NO. 17) domains than was the case for  $\beta$ -lactamase;  $\beta$ -lactamase could be strongly destabilized *in vitro* by fusion with as few as two ubiquitinG76V domains (SEQ. ID. NO. 17) (**FIG. 2A**) whereas GFP required at least three ubiquitinG76V domains (SEQ. ID. NO. 17) to be strongly destabilized (**FIG. 3**). This relationship between the destabilization domain, and the protein to be destabilized, emphasizes the utility of the multiubiquitin destabilization system, in that the extent of destabilization can be manipulated to give the desired properties by altering the number of ubiquitinG76V (SEQ. ID. NO. 17) domains that are present in the destabilization domain.

#### **Example 7. Characterization of multimerized destabilization domain-endogenous mammalian protein fusions *in vitro*.**

Characterization of the turnover of multiubiquitin-caspase-3 fusion proteins *in vitro* was performed as described in Example 5. The TNT synthesis reactions were diluted into chase lysate in the presence of cycloheximide and chase time points were taken and analyzed by SDS-PAGE and autoradiography. **FIG. 4** shows that wild-type caspase-3 is stable over a 60 minute chase *in vitro*, and that fusion to the multiubiquitin destabilization domain results in rapid degradation. In particular, the ubiquitinG76V-caspase-3 fusions are degraded in a very similar manner to the ubiquitinG76V- $\beta$ -lactamase fusions although the Ub-caspase-3 fusions appear to be degraded slightly slower *in vitro* than the Ub- $\beta$ -lactamase fusions. Altogether, these data demonstrate the generalized applicability of the multiubiquitin destabilization domain approach to provide for predictable destabilized of any given chosen target protein using this system.

**Example 8. Characterization of the half-life of multimerized destabilization domain- $\beta$ -lactamase fusion proteins within cells.**

UbiquitinG76V- $\beta$ -lactamase constructs in pcDNA3 (SEQ. ID. NOs. 23 to 26) were introduced into Jurkat T-lymphocytes by electroporation. Stable transfectants were selected in RPMI 1640 media containing 10% fetal bovine serum (Gibco) and 0.8 mg/ml G418 (Geneticin, Gibco). Analysis of  $\beta$ -lactamase activity in intact Jurkat cells stably transfected with the pcDNA3-1-4XUb-Bla (SEQ. ID. NOs. 23 to 26) constructs was accomplished by loading the cells with the fluorescent  $\beta$ -lactamase substrate CCF2/AM as described in Zlokarnik *et al.* (1998) (Science 279, 1848) followed by analysis by fluorescence activated cell sorter (Becton Dickinson FACSTM VantageTM) or CytoFluor microtiter plate fluorimeter (Perseptive Biosystems). For kinetic measurements, to determine the half-life of the fusion protein *in vivo*, direct measurements were made of  $\beta$ -lactamase activity in lysates prepared from cells expressing the various ubiquitinG76V-Bla fusions.

Flow cytometry and cell sorting were conducted using a Becton Dickinson FACSTM VantageTM with a Coherent Enterprise II<sup>TM</sup> argon laser producing 60mW of 351-364 nm multi-line UV excitation. The flow cytometer was equipped with pulse processing and the Macrosort<sup>TM</sup> flow cell. Cells were loaded with 1  $\mu$ M CCF2/AM for 1-2 hours at room temperature prior to sorting, and fluorescence emission was detected via 460/50nm (blue) and 535/40nm (green) emission filters, separated by a 490nm long-pass dichroic mirror. The results from one such experiment are shown in **FIG. 5**, where the abundance of cells expressing relatively high levels of  $\beta$ -lactamase (regions R5+R6+R7) was determined. This analysis showed that the relative abundance of cells expressing high steady state levels of  $\beta$ -lactamase was inversely proportional to the number of copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to  $\beta$ -lactamase, i.e., the lowest levels of  $\beta$ -lactamase expression were found in cells expressing  $\beta$ -lactamase fusions containing the most copies of ubiquitinG76V (SEQ. ID. NO. 17).

Similar cytometric analysis experiments were used to investigate the degradation properties of multiUb-Bla fusions *in vivo*. Jurkat cells expressing multiUb-Bla fusions were treated with 50  $\mu$ M MG132 to investigate whether the low

5  $\beta$ -lactamase activity found in cells expressing 3-4XUb-Bla requires proteasome activity. The results, shown in **Table 5**, below show that the addition of inhibitor (+inh/-chx samples) results in a significant increase in the percentage of positive BLA expressing cells for the 2X, 3X and 4X ubiquitinG76V fusion protein constructs compared to the untreated controls (-inh/-chx samples.)

TABLE 5				
	-inh/-chx %Bla <sup>+</sup> cells	+inh/-chx %Bla <sup>+</sup> cells	-inh/+chx %Bla <sup>+</sup> cells	+inh/+chx %Bla <sup>+</sup> cells
WT Bla	22.5	22.7	17.6	19.0
1XUb-Bla	17.4	18.8	8.5	16.2
2XUb-Bla	12.0	17.1	2.1	12.2
3XUb-Bla	8.3	14.6	1.5	9.8
4XUb-Bla	4.1	12.1	0.5	5.0

10 Furthermore, treating these cells with 100  $\mu$ g/ml cycloheximide (to block protein synthesis) for one hour prior to CCF2 loading and cytometric analysis (compare columns [-inh/+chx] and [-inh/-chx]) resulted in a strong decrease in  $\beta$ -lactamase activity only in cells expressing 2-4XUb-Bla and this decrease could largely be blocked by preincubating the cells with 50  $\mu$ M MG132 prior to cycloheximide addition (column +inh/+chx, in **Table 5**).

15 These data are strong evidence that the multiubiquitin domain in ubiquitinG76V-Bla fusions is acting as a destabilization motif that directs the rapid degradation of the fusions in a proteasome-dependent manner that is controlled by the number of ubiquitinG76V (SEQ. ID. NO. 17) domains within the multiubiquitin destabilization domain.

20 In order to determine accurate quantitative measurements of the kinetic characteristics of the degradation of UbG76V- $\beta$ -lactamase fusions *in vivo*,  $\beta$ -lactamase activity was determined in cellular lysates. To do this, Jurkat cells expressing the various forms of multiUb-Bla fusion proteins were sorted by flow cytometry to obtain a pool of cells representative of the Bla<sup>+</sup> population seen in **FIG.**  
 25 **5** (Region R5+R6+R7). These cells were treated with 100  $\mu$ g/ml cycloheximide to inhibit new protein synthesis, and aliquots of cells were taken at appropriate intervals,

to measure the  $\beta$ -lactamase activity remaining. This approach enabled a determination of the rate of destruction of the cellular pool of  $\beta$ -lactamase fusion proteins within the cell.  $\beta$ -lactamase activity was determined in these cell samples by transferring them to ice to terminate further metabolism, and then pelleted by centrifugation. The cell pellets were converted to lysates and  $\beta$ -lactamase activity was measured *in vitro* using the free acid form of the  $\beta$ -lactamase substrate CCF2. Aliquots of the lysates were assayed using 10  $\mu$ M CCF2 in PBS at room temperature. Hydrolysis of the fluorescent substrate was monitored in a Perseptive Biosystems CytoFluor plate reader using a 395/25nm excitation filter and 460/40nm emission filter.

In agreement with the cell analyses by flow cytometry, cells expressing wild-type  $\beta$ -lactamase had high levels of  $\beta$ -lactamase activity, that was relatively resistant to proteolytic degradation over a 90 minute incubation period with cycloheximide; wild-type  $\beta$ -lactamase activity decayed with a half-life >2 hours (**FIG. 6**). Cells expressing 1XUb-Bla fusions also contained relatively high levels of  $\beta$ -lactamase activity that decayed with a half-life of about 20-30 minutes. Cells expressing  $\beta$ -lactamase fused to 2 or more copies of ubiquitinG76V (SEQ. ID. NO. 17) had significantly less  $\beta$ -lactamase activity at steady state (compare 0 minute time points) and the half-lives of these pools of fusion proteins were strikingly short, with all three fusion proteins decaying with *in vivo* half-lives of less than 10 minutes.

The  $\beta$ -lactamase measurements from the Jurkat cell lysates allows a calculation of the intracellular concentration and copy number of  $\beta$ -lactamase fusion proteins in the respective cell lines. A standard curve created of the hydrolysis of CCF2 by purified  $\beta$ -lactamase enzyme was generated and used to calculate the steady state concentration of  $\beta$ -lactamase fusion protein for each cell line. This analysis showed that there was a ten-fold difference in intracellular concentration between wild-type  $\beta$ -lactamase and 4XUb- $\beta$ -lactamase at steady state (**Table 6**). The calculated concentration of wild-type  $\beta$ -lactamase corresponds to 21,000 molecules per cell, in very good agreement with the values reported by Zlokarnik *et al.* (1998) (*Science* 279, 1848) for cells expressing high levels of wild-type  $\beta$ -lactamase.

TABLE 6		
Construct	Half-life	Intracellular Concentration
WT Bla	>120 min	35 nM
1XUb-Bla	20-30 min	30 nM
2XUb-Bla	<10 min	7 nM
3XUb-Bla	<10 min	5 nM
4XUb-Bla	<10 min	3.5 nM

The kinetic data on fusion protein turnover, together with the steady state concentration measurements, demonstrate that the fusion of a multiubiquitin destabilization domain to a target protein allows for the manipulation of both the intracellular concentration, as well as, the turnover kinetics of the resulting fusion proteins. The present invention provides for a method of regulating the intracellular concentration of any target protein within a cell, independently of the rate of transcription of that protein. Unlike other systems of regulating the intracellular concentrations of target proteins, the present invention provides for the ability to "preset" the final concentration of the target protein within a ten-fold range of expression.

The data with multiubiquitinG76V- $\beta$ -lactamase fusions demonstrate that fusions containing one to four copies of ubiquitinG76V fused to  $\beta$ -lactamase results in chimeric proteins with half-lives *in vivo* of from 5 to 30 minutes. There are likely to be applications that require proteins that have a half-life longer than that obtained with fusion to one copy of ubiquitinG76V. For such instances, it would be useful to have a form of uncleavable ubiquitin that is recognized by E2/E3 ubiquitin ligases with lower affinity and therefore result in less destabilization than with fusions to ubiquitinG76V. The efficient recognition and degradation of proteins by the proteasome requires the formation of extended polyubiquitin chains that are extended in isopeptide linkage between a critical lysine residue on ubiquitin to the C-terminus of the incoming ubiquitin. The internal lysine in ubiquitin most often used in such polyubiquitin chains is lysine-48. In order to create a longer half-life protein, it is

recognized that it is possible to mutagenize the ubiquitin homolog fused to the protein of interest such that it is not recognized by E2/E3 ubiquitin ligases as efficiently as wild-type ubiquitin. It is likely that mutagenesis of lysine-48, (to Arg, His, Gln or Asn for example) and / or the residues surrounding it will yield a form of ubiquitin that is recognized and extended with lower affinity, than the non-mutant forms. The non extendable homologs would thus serve to create fusion proteins with longer half lives than is otherwise possible with wild-type ubiquitin. Typically such constructs would contain between one and five copies of the non-extendable, non-cleavable ubiquitin homologs to provide for a wide range of destabilization.

Alternatively, random mutagenesis of the ubiquitin or mutation of other lysines in ubiquitin may result in a form of ubiquitin with the desired properties.

#### **Example 9. Characterization of the stability of multimerized destabilization domain-Naturally Fluorescent Protein fusions within cells.**

UbiquitinG76V-GFP constructs in pcDNA3 (SEQ. ID. NOs. 31 to 34) were introduced into CHO cells by Lipofectamine (Life Technologies) transfection. Stable transfectants were selected in RPMI 1640 media containing 10% fetal bovine serum (Gibco) and 0.8 mg/ml G418 (Geneticin, Gibco). Analysis of GFP fluorescence in CHO cells stably transfected with various ubiquitinG76V-GFP constructs was analyzed by flow cytometry on a Becton Dickinson FACST<sup>™</sup> Vantage<sup>™</sup> with a Coherent Enterprise II<sup>™</sup> argon laser producing 60mW of 488nm UV excitation. The flow cytometer was equipped with pulse processing and the Macrosort<sup>™</sup> flow cell. Fluorescence emission was detected via 530/30nm emission filter. The FACS analyses of stable populations determined that the steady state percentage of bright green GFP<sup>+</sup> cells varied depending on the presence of the multiubiquitin destabilization domain. The relative percentages of GFP<sup>+</sup> cells are shown in the **Table 7**.

<b>TABLE 7</b>	
<b>Stable CHO cell line</b>	<b>% GFP<sup>+</sup> cells</b>
Wild-type GFP	39.13
1XUb-GFP	5.74

2XUb-GFP	3.06
3XUb-GFP	2.2
4XUb-GFP	1.93

This analysis showed that the relative abundance of cells expressing high steady state levels of GFP fluorescence was inversely proportional to the number of copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to the protein, i.e., the lowest levels of GFP-expressing cells were found in the fusions containing the most copies of ubiquitinG76V (SEQ. ID. NO. 17). The steady state concentration measurements demonstrate that fusions of a multiubiquitin destabilization domain to the highly stable GFP mutant Emerald (SEQ. ID. NO. 28) allows for the predictable and controllable manipulation of the intracellular concentrations of naturally fluorescent proteins.

#### **Example 10. Construction of destabilization domain – linker – reporter moiety fusion proteins**

Ubiquitin- $\beta$ -lactamase fusion proteins containing a specific protease cleavage site were constructed by annealing the complementary oligonucleotides DEVD-1 (SEQ. ID. NO. 45) and DEVD-2 (SEQ. ID. NO. 46) that encode a caspase-3-type cleavage site and produce BamH I compatible ends. This oligonucleotide cassette was ligated into BamH I-digested pcDNA3-1-4XUb-Bla plasmid constructs (SEQ. ID. NOS. 23 to 26) described in Example 2. The resulting constructs encode an in-frame fusion protein consisting of from one, to four, copies of ubiquitinG76V (SEQ. ID. NO. 17) separated from  $\beta$ -lactamase by linker containing a caspase-3 cleavage site; the plasmids were designated as pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOS. 47-50). A control linker containing a DEVA site that should not serve as a cleavage site for caspase-3-like proteases was constructed in an identical manner using DEVA1 (SEQ. ID. NO. 51) and DEVA2 primers (SEQ. ID. NO. 52) and the resulting plasmids were designated as pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. Nos. 53-56).



↓ ↓

**LVLRLRGVSGVAVGSGVDEVDGSGAWLHPETLVKV**

**Example 11. Detection of caspase activity using destabilized reporter moieties *in vitro***

<sup>35</sup>S-labeled ubiquitin-β-lactamase fusion proteins containing a cleavage site for the group II effector caspase-3 were produced by *in vitro* transcription/translation reactions as described in Example 2 except that plasmids pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOs. 47-50) or control plasmids pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. NOs. 53-56) were used as templates. The <sup>35</sup>S-labeled proteins were then used as substrates for purified caspase-3 in an *in vitro* cleavage reaction. The 12 μl reaction consisted of 4 μl of <sup>35</sup>S-labeled ubiquitin-DEVD/A-Bla fusion proteins, 100 mM HEPES pH 7.5, 10% sucrose, 0.1% CHAPS, 10 mM DTT and 25 nM purified recombinant caspase-3. The reactions were incubated at 30°C and samples taken at 0, 5, 10, 20, 30, 45, and 60 minutes and analyzed by SDS-PAGE and autoradiography. The results from 2XUb-DEVD-Bla and 2XUb-DEVA-Bla fusion proteins are shown in **FIG. 7A**. The 2XUb-DEVD-Bla fusion served as a very good substrate for caspase-3 with over 90% cleavage within 5 minutes. In contrast, the 2XUb-DEVA-Bla fusion was not cleaved by caspase-3 *in vitro*, even at extended incubation times. The 2XUb-DEVD-Bla cleavage product seen in **FIG. 7A** co-migrates on SDS-PAGE gels with β-lactamase fused to the short DEVD linker region (data not shown) and verifies the position of the cleavage site and identifies the labeled cleavage product as the β-lactamase portion of the cleaved fusion. The liberated destabilization domain is much smaller and has run off the gel in this experiment. These data demonstrate that the DEVD fusion serves as an efficient substrate for caspase-3 and the lack of cleavage with the DEVA fusion confirms that the cleavage is occurring at the DEVD site.

The protease assay outlined above requires that the protease cleavage result in a stabilization of the catalytic domain of the reporter. To test whether this is the case,

we mixed approximately equal portions of cleaved and uncleaved <sup>35</sup>S-labeled reporters from *in vitro* cleavage reactions identical to those in FIG. 7A and then diluted the fragments into crude chase lysate containing cycloheximide to perform a chase experiment. The reactions were incubated at 37°C and samples were taken at 0, 5, 10, 20, 30 and 60 minutes and analyzed by SDS-PAGE and autoradiography. FIG. 7B shows that the uncleaved intact 2XUb-DEVD-Bla or 2XUb-DEVA-Bla reporters were degraded very rapidly *in vitro* with a half-life of less than 5 minutes. In contrast, the cleavage product from the 2XUb-DEVD-Bla reporter lacks the destabilization domain and as a result is very stable *in vitro*. These data confirm that the intact and cleaved versions of the β-lactamase reporters have dramatically different half-lives and provide evidence that this difference in stability may provide a format for assaying endoprotease activity *in vivo*.

#### Example 12. Detection of effector caspase protease activity using destabilized reporter moieties within cells

Plasmids pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOs. 47-50) and pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. NOs. 53-56) were transfected into Jurkat cells and selected for stable transfectants as described in Example 8. The stable transfectants were sorted by flow cytometry using Becton Dickinson FACS™ Vantage™ SE and FACS™ Vantage™ flow cytometers. The FACS™ Vantage™ SE was equipped with Turbosort Option, pulse processing, ACDU, and Coherent Innova 302C krypton and Coherent Innova 70 Spectrum mixed-gas krypton-argon lasers. The FACS™ Vantage™ was equipped with pulse processing, ACDU, and Coherent Enterprise II and Coherent Innova 70 Spectrum mixed-gas krypton-argon (with violet option) lasers. For β-lactamase experiments, 60mW of 413nm laser emission was used for CCF2 excitation, with a 500nm dichroic filter separating a 460/50nm (CCF2 blue fluorescence) and a 535/40nm bandpass filter (green fluorescence). Single cells with the desired level of β-lactamase expression were sorted into individual wells of 96-well plates using the Automatic Cell Deposition Unit (ACDU) on the FACS™ Vantage™ and expanded for analysis of homogeneous clonal populations. All results in this Example utilized clonal lines.

The clonal cell lines were initially screened for expression of  $\beta$ -lactamase and the ability to degrade the Ub-DEVD-Bla or Ub-DEVA-Bla fusion rapidly. This initial screen was accomplished by treating an aliquot of cells with 100  $\mu$ g/ml cycloheximide followed by incubation at 37°C for 1 hour (chase period). Treated and untreated cells were loaded with 1  $\mu$ M CCF2-AM for 1 hour at room temperature and  $\beta$ -lactamase levels were quantified using a CytoFluor microtiter plate fluorimeter (Perseptive Biosystems) using 395/25nm excitation and 460/40 (blue) nm and 530/30 (green) nm emission filters. Emission ratios were calculated from background-subtracted values (background=media+CCF2 alone) and expressed as a 460/530nm ratio where a high ratio indicates high  $\beta$ -lactamase activity. This analysis showed that Ub-DEVD-Bla fusions with two or more copies of ubiquitinG76V (SEQ. ID. NO. 17) gave satisfactory chase characteristics, with fusions to two copies of UbiquitinG76V (SEQ. ID. NO. 17) giving the highest steady state levels (no chase) of fusion protein (data not shown). In contrast, 1XUb-DEVD-Bla fusions were not sufficiently destabilized to be usable with this assay format as cells expressing the fusion required extended cycloheximide treatments (data not shown). As the 2-4XUb-DEVD-Bla fusions all exhibited satisfactory rates of proteolytic turnover in cells, the 2X ubiquitinG76V destabilization domain was used with the DEVD-Bla fusions because it gave the best performance (expression levels vs. turnover kinetics) in this particular application. It is worth noting here that due to the variability in the intrinsic stability of different proteins fused to the ubiquitinG76V (SEQ. ID. NO. 17) destabilization domain; fusions of other cellular proteins with multimerized destabilization constructs would be expected to require a dissimilar number of copies of ubiquitinG76V (SEQ. ID. NO. 17) to impart sufficiently rapid turnover kinetics (data not shown). A key advantage of the present invention is the ability to meet this need by varying the number of destabilization domains present within the multimerized destabilization domain construct.

One clonal cell line from each of 2XUb-DEVD-Bla and 2XUb-DEVA-Bla cell populations was characterized in detail. To establish the background (no  $\beta$ -lactamase) control value, wild-type Jurkat cells containing no  $\beta$ -lactamase activity were loaded with CCF2-AM and the 460/530 fluorescence ratio measured. The value

obtained, about 0.05, establishes the background ratio exhibited by cells in the absence of  $\beta$ -lactamase activity. When the 2XUb-DEVD-Bla and 2XUb-DEVA-Bla clones were treated with cycloheximide (chx) for 1 hour at 37°C prior to CCF2-AM loading, they both exhibited 460/530 ratios very near the background ratio of 0.05, demonstrating that the cells retained the ability to degrade the 2XUb-Bla fusion very efficiently (**Table 8**).

<b>TABLE 8</b>		
	<b>2XUb-DEVD-Bla</b> 460/530 emission ratio	<b>2XUb-DEVA-Bla</b> 460/530 emission ratio
no chx	1.80	1.60
+ chx	0.07	0.07
+ $\alpha$ Fas/-chx	1.25	1.10
+ $\alpha$ Fas/+chx	0.67	0.12
+ $\alpha$ Fas/+inh/+chx	0.08	0.09

The fact that there is a significant difference in stability between the uncleaved reporter and the cleavage product *in vitro* (**FIG. 7B**) forms the basis for an assay for protease activity in intact cells. As shown in **Table 8**, in the absence of caspase activity, both 2XUb-DEVD-Bla and 2X-Ub-DEVA-Bla fusions are rapidly degraded to very low levels in the presence of cycloheximide to inhibit new protein synthesis. Treatment of Jurkat cells with Fas ligand will result in the activation of Fas receptor - an apoptosis signaling receptor found on the surface of a number of cell types that belongs to the tumor necrosis factor (TNF)/nerve growth factor family. Fas activation ultimately leads to the activation of the group II caspases that efficiently cleave substrates containing DEVD recognition motifs. In order to activate this pathway and measure the activity of group II caspases using the DEVD-Bla reporter in intact cells, an anti-Fas antibody (CH-11 anti-Fas IgM; Kamiya Biomedical Co., Seattle, WA) was used to cross-link the receptor and stimulate the activation of group II caspases. Western blot analysis of the anti-Fas-treated cells confirmed the proteolytic activation of caspase-3 (data not shown), the major group II caspase activity in Jurkat cells. Treatment of Jurkat cells expressing 2XUb-DEVD-Bla or 2XUb-DEVA-Bla reporter with 50 ng/ml anti-Fas IgM for 6 hours at 37°C resulted in a modest decrease in the steady-state levels of the reporter (**Table 8**), most likely due to the inhibition of

protein synthesis that is known to accompany apoptosis. At this point, the activation of group II caspases will result in the cleavage and stabilization of some proportion of the DEVD-Bla (but not the control DEVA-Bla) reporters. Treatment of such cells with cycloheximide would then allow for the clearing of the uncleaved, short half-life reporters while leaving the stable cleaved reporters as the sole forms of  $\beta$ -lactamase activity in the cells. **Table 8** shows that cycloheximide addition to anti-Fas treated cells (+ $\alpha$ Fas/+chx) resulted in the stabilization of a significant fraction of the DEVD-Bla reporters while the DEVA-Bla reporters cannot be cleaved and stabilized. To show that the stabilization of the DEVD-Bla reporters is due to caspase activation, we used the peptide inhibitor Z-VAD-fmk (Enzyme Systems Products, Livermore, CA) that is a potent broad inhibitor of caspases. Treatment of the cells with 10  $\mu$ M Z-VAD-fmk coincident with anti-Fas addition blocked the stabilization of DEVD-Bla reporters. Treatment of the cells with cycloheximide resulted in the degradation of the non cleaved constructs to background levels of  $\beta$ -lactamase activity (+ $\alpha$ Fas/+Inh/+chx). Comparison of  $\beta$ -lactamase levels in antiFas-treated DEVD-Bla-expressing cells in the presence or absence of Z-VAD-fmk inhibitor determines the dynamic range of the assay; in this particular experiment the dynamic range is approximately 8-fold. These data demonstrate that the cleavage and stabilization of short half-life  $\beta$ -lactamase protease reporters provides a sensitive and specific assay for measuring the activation of caspases in intact cells.

It is of note that this assay format would permit the identification of compounds that stimulate group II caspases and subsequent apoptosis (agonist/inducer format) as well as compounds that inhibit caspase activity stimulated by a known reagent such as anti-Fas IgM (antagonist/inhibitor format). As evidence for this assay being useful for both inducer and inhibitor applications, we generated dose-response curves for both an inducer of caspases and apoptosis (anti-Fas IgM) and an inhibitor of anti-Fas induced apoptosis (Z-VAD-fmk). **FIG. 8** shows that the assay in Jurkat cells expressing 2XUb-DEVD-Bla generates sufficient dynamic range to detect low concentrations of the inducer anti-Fas IgM ( $EC_{50}$ = 11 ng/ml). In addition, treatment of Jurkat cells expressing 2XUb-DEVD-Bla with 50 ng/ml anti-

Fas IgM allows sensitive detection of inhibition by Z-VAD-fmk with  $IC_{50}= 5 \mu M$  (FIG. 8).

5 **Example 13. Creation of reporters for viral self-cleaving proteases using multimerized destabilization domain- $\beta$ -lactamase-rhinovirus 2A protease fusions.**

The gene encoding the human rhinovirus 14 2A protease (SEQ. ID. NO. 57) was isolated by PCR amplification from genomic RNA by RT-PCR using  
10 oligonucleotides HRV145' (SEQ. ID. NO. 58) and HRV143', (SEQ. ID. NO. 59). The resulting PCR product had BamH I sites at both ends of the HRV14 2A protease sequence and could be inserted in frame into the pcDNA3-1-4XUb-Bla vectors (SEQ. ID. Nos. 23-26) from example 2. The PCR fragment from this reaction was digested with BamH I and ligated into pcDNA3-3XUb-Bla (SEQ. ID. NO. 25). The resulting  
15 construct, pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60) was further characterized *in vitro* and within cells.

In addition to the HRV14 2A protease constructs, two additional constructs were made for the HRV16 2A protease. The gene for the human rhinovirus 16 sequence 2A protease (SEQ. ID. NO. 61) was isolated by polymerase chain reaction  
20 (PCR) amplification of a plasmid template. The PCR template was a plasmid construct containing the entire HRV16 genome (a gift from Dr. Wai Ming Lee at the University of Wisconsin). Oligonucleotides HRV165' (SEQ. ID. NO. 62) and HRV163', (SEQ. ID. NO. 63) were used in a PCR reaction with the HRV16 plasmid resulting in a PCR product that had BamH I sites at both ends of the HRV16 2A  
25 protease sequence. The PCR fragment from this reaction was digested with BamH I and ligated into pcDNA3-3XUb-Bla (SEQ. ID. NO. 25) and pcDNA3-Ub-Met Bla (SEQ. ID. NO. 27) via the BamH I site. The resulting constructs were pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO. 64) and pcDNA3-Ub-Met-Bla HRV16 (SEQ. ID. NO. 65). In addition, two mutant constructs were made for the HRV16 2A protease.  
30 These mutants corresponded to mutations at two residues of the putative catalytic triad for the 2A protease and should result in a catalytically inactive mutant, specifically, aspartate 35 was mutated to alanine (D35A) and cysteine 106 was mutated to alanine (C106A). These derivatives were generated by mutagenesis of the

HRV16 2A protease using oligonucleotide HRV16 D35A (SEQ. ID. NO. 66) and oligonucleotide HRV16 C106A (SEQ. ID. NO. 67). The resulting plasmids were designated as pcDNA3-3XUb-Bla HRV16(C106A) (SEQ. ID. NO. 68), pcDNA3-3XUb-Bla HRV16(D35A) (SEQ. ID. NO. 69), pcDNA3-Ub-Met-Bla HRV16(C106A) (SEQ. ID. NO. 70) and pcDNA3-Ub-Met-Bla HRV16(D35A) (SEQ. ID. NO. 71).

**Example 14. Detection of Rhinovirus protease activity using destabilized reporter moieties in vitro.**

<sup>35</sup>S-labeled ubiquitin- $\beta$ -lactamase fusion proteins containing the HRV14 and HRV16 2A proteases, as well as the mutants above, were produced by *in vitro* transcription/translation reactions as described in Example 5. The plasmids pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO. 64), pcDNA3-3XUb-Bla HRV16(C106A) (SEQ. ID. NO. 68), pcDNA3-3XUb-Bla HRV16(D35A) (SEQ. ID. NO. 69), pcDNA3-Met Ub-Bla HRV16 (SEQ. ID. NO. 65), pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60), and pcDNA3-MetUb-Bla HRV14 (SEQ. ID. NO. 72) were used as templates. The reactions were incubated at 30°C for 45 min and analyzed by SDS-PAGE and autoradiography. **FIG. 9A** shows the results of TNT synthesis reactions for the wild-type HRV16 2A and the two mutant HRV16 2A constructs. Shown are the levels of expression for the stable (Met) and destabilized 3X ubiquitinG76V HRV16 2A-Bla fusions. As expected, the level of expression is higher in the stable methionine containing constructs than the destabilized 3XUb constructs (**FIG. 9A**). The wild-type HRV16 2A fusions also show significant accumulation of the lower molecular weight stable cleavage product indicating that the fusions exhibit robust autocatalytic cleavage activity in these *in vitro* reactions. In contrast, mutation of residues in the putative catalytic triad (aspartate 35 and cysteine 106) blocked formation of the stable cleavage product, indicating that these mutants are indeed catalytically inactive.

The protease assay outlined in Example 10 requires that protease cleavage results in a stabilization of the catalytic domain of the reporter. To test for this requirement the pcDNA3-3XUb-Bla HRV14 TNT reaction was diluted into chase lysate containing cycloheximide to perform a chase experiment. The reactions were incubated at 37°C for 60 minutes and analyzed by SDS-PAGE and autoradiography.

**FIG. 9B** shows that the uncleaved intact 3XUb-HRV14-Bla reporter was completely degraded during the 60 minute chase. In contrast, the cleavage product from the 3XUb-HRV14-Bla reporter lacks the destabilization domain, and as a result, is stable *in vitro*. These data confirm that the intact and cleaved versions of the HRV 2A- $\beta$ -lactamase fusion reporters have dramatically different half-lives and provide evidence that this difference in stability can provide the basis for assaying self-cleaving protease activity in side intact cells.

**Example 15. Detection of Rhinovirus protease activity using destabilized reporter moieties *in vivo*.**

The biochemical properties of self-cleaving *cis* proteases such as rhinovirus 2A pose several technical challenges that have hampered the development of a screening format to allow for the identification of inhibitors or activators in cell based assays. First, the activity of the protease is directed toward cleavage of itself. This rules out the use of separate reporters that are cleaved in *trans* and limit the catalytic output of the assay, i.e., a single protease molecule generates a single cleavage product and this fact eliminates the catalytic amplification used in traditional assays for trans-cleaving proteases. In order to address these limitations, the  $\beta$ -lactamase reporters are incorporated into the 2A protease itself, thereby measuring the *cis* cleavage reaction directly and gaining the advantage of a catalytic reporter that can cleave many CCF2 substrate molecules per reporter. Since the HRV 2A protease undergoes the self-cleavage reaction immediately upon synthesis, the screening assay must be performed on newly synthesized HRV 2A- $\beta$ -lactamase reporters. A screen to identify inhibitors of the protease must incorporate a step where test compounds are added and their effect then measured. As cleaved stable  $\beta$ -lactamase reporters will accumulate in the cell as the HRV 2A-Bla reporters are being constitutively expressed, it is essential to eliminate the readout due to such cleavage products that are generated before the test compound is added. To do this, the  $\beta$ -lactamase inhibitor clavulanate was used. Clavulanate is a non-cytotoxic irreversible inhibitor of  $\beta$ -lactamase and overnight treatment of Jurkat cells reduces  $\beta$ -lactamase levels to background (See commonly owned U.S. Patent Application No. 09/067,612 filed



April 28 1998). Therefore, clavulanate treatment of Jurkat cells expressing HRV 2A-Bla fusions eliminates the  $\beta$ -lactamase activity that is present in the cell resulting from both uncleaved and cleaved  $\beta$ -lactamase reporters. In essence, this has the effect of "zeroing out" the  $\beta$ -lactamase activity in the cells and bringing the cells back down to baseline activity. The clavulanate can then be washed out and test compound added. New synthesis of HRV 2A-Bla reporters will result in the accumulation of the fusion protein reporter in the cells and the self-cleavage reaction will now be subject to inhibition by the test compound. After an appropriate interval to allow for the cleavage of newly synthesized reporters has passed, the cells can be treated with cycloheximide to clear out the unstable uncleaved reporters and the resulting  $\beta$ -lactamase activity will be due exclusively to cleaved, stabilized reporters.

Plasmids pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO. 64) and pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60) were transfected into Jurkat cells and selected for stable transfectants as described in Example 8. The stable transfectants were sorted by flow cytometry using Becton Dickinson FACST<sup>TM</sup> Vantage<sup>TM</sup> SE and FACST<sup>TM</sup> Vantage<sup>TM</sup> flow cytometers. The FACST<sup>TM</sup> Vantage<sup>TM</sup> SE was equipped with Turbosort Option, pulse processing, and Coherent Innova 302C krypton and Coherent Innova 70 Spectrum mixed-gas krypton-argon lasers. The FACST<sup>TM</sup> Vantage<sup>TM</sup> was equipped with pulse processing, and Coherent Enterprise II and Coherent Innova 70 Spectrum mixed-gas krypton-argon (with violet option) lasers. For  $\beta$ -lactamase experiments, 60mW of 413nm laser emission was used for CCF2 excitation, with a 500nm dichroic filter separating a 460/50nm (CCF2 blue fluorescence) and a 535/40nm bandpass filter (CCF2 green fluorescence). Single cells with the desired level of  $\beta$ -lactamase expression were sorted into individual wells of 96-well plates using the Automatic Cell Deposition Unit (ACDU) on the FACST<sup>TM</sup> Vantage<sup>TM</sup> and expanded for analysis as homogeneous clonal populations. All results in Example 15 utilized clonal lines.

Selected clones (25-50 for each construct) were then expanded further for analysis. Clones were treated for 16 hours with 300  $\mu$ M clavulanate, washed twice with phosphate buffered saline (PBS), incubated for 2 hours at 37°C, treated for 1 hour at 37°C with 100  $\mu$ g/ml cycloheximide, and then loaded with CCF2-AM for 2

hours at room temperature. The individual clones were then screened visually by fluorescence microscopy. At least 24 individual clones were tested in this manner for each construct and one clone chosen for each construct.

To assay HRV 2A protease activity, the selected Jurkat stable cell clones were  
5 treated for 16 hours with 300  $\mu$ M clavulanate to inactivate pre-existing cleaved and  
uncleaved HRV 2A-Bla fusion protein. Cells were then washed twice with PBS,  
resuspended at 100,000 cells/well in 100  $\mu$ l RPMI + 10% FBS in 96-well plates. The  
cells were incubated at 37°C for 4 hours in the presence or absence of an inhibitor of  
the 2A protease. Cells were treated with 100  $\mu$ g/ml cycloheximide for 30 minutes at  
10 37°C, loaded with CCF2-AM for 2 hours at room temperature and read on the  
CytoFluor plate reader as described in Example 8. Inhibitor compounds, radicicol and  
geldanamycin, were used for the validation of the HRV protease cell-based assay.  
These compounds are known inhibitors of the Hsp90 heat shock protein (see Roe et  
al., (1999) J. Med. Chem. 42 260-266), which is required for the folding and  
15 regulation of a number of cellular proteins and can inhibit HRV 2A protease activity  
*in vitro* (data not shown). Compounds were tested at 1  $\mu$ M for their ability to inhibit  
the HRV 2A protease cell-based assay using clones expressing HRV16 and HRV14  
2A protease reporters. Jurkat cells expressing 3XUb-Bla-HRV14 or HRV16 2A  
protease fusion proteins contained significant  $\beta$ -lactamase activity in the absence of  
20 the inhibitors (**Table 9**). Both radicicol and geldanamycin showed strong inhibition  
of cellular  $\beta$ -lactamase activity remaining after the cycloheximide chase. The  
inhibitors are not simply inhibiting  $\beta$ -lactamase enzyme activity because control  
experiments showed that radicicol and geldanamycin did not inhibit  $\beta$ -lactamase  
activity in Jurkat cells expressing wild-type  $\beta$ -lactamase (data not shown). These data  
25 demonstrate that the  $\beta$ -lactamase activity present after a cycloheximide chase is due to  
HRV 2A protease activity and that this  $\beta$ -lactamase activity can be blocked using  
inhibitors of HRV 2A protease. These results further demonstrate that Jurkat cells  
expressing 3XUb-Bla HRV 2A fusion proteins constitute a robust cell-based assay for  
HRV 2A cis-protease activity. The difference in  $\beta$ -lactamase activity between  
30 untreated and inhibitor-treated cells determines the dynamic range of this assay; in  
this particular experiment, the assay dynamic range is approximately 6-fold.

TABLE 9		
	3XUb-HRV14-Bla 460/530nm ratio	3XUb-HRV16-Bla 460/530nm ratio
no inhibitor	1.022	0.895
+ radicicol	0.152	0.229
+ geldanamycin	0.153	0.239

**Example 16. Detection of Proteasome activity within cells using destabilized reporter moieties and use in the identification of proteasome inhibitors.**

5 A direct application of the destabilized reporter fusions is in the measurement of the activity of the proteolytic activity that responsible for the constitutive degradation of the reporter in cells. Ubiquitinated proteins are known to be degraded by the multi-subunit proteasome. In addition, the proteasome is responsible for the degradation of the large majority of cellular proteins see Lee and Goldberg, (1998)

10 Trends Cell Biol., 8 397-403). The proteasome itself has been implicated in a number of pathological conditions resulting from either increased or decreased proteasome activity (see Ciechanover, (1998) EMBO J. 17 7151-7160). As such, the proteasome represents an attractive target for intervention in pathological conditions using small molecule inhibitors or activators.

15 Inhibitors of the proteasome were initially tested *in vitro* for inhibition of degradation of 2XUb-Bla. Transcription/translation reactions on the pcDNA3-2XUb-Bla (SEQ. ID. NO. 24) construct were preformed as described in Example 5. The <sup>35</sup>S-labeled synthesis reactions were diluted into crude chase lysates in the presence of cycloheximide and inhibitor and incubated at 37°C for 20 minutes. Samples were

20 then analyzed by SDS-PAGE and autoradiography. **FIG. 10** shows that >90% of the starting <sup>35</sup>S-labeled fusion protein is degraded by the 20 minute time point in the absence of proteasome inhibitors. Addition of the inhibitor MG132 (Calbiochem) at 50 μM resulted in a significant increase in the intact, un-conjugated fusion protein as well as the appearance of high molecular weight labeled species that represent

25 extensive further ubiquitination of the fusion protein. The high molecular weight ubiquitin conjugates accumulate prominently in the presence of MG132 because they are recognized so efficiently by the proteasome that they are barely visible without

inhibiting their degradation. Additional proteasome inhibitors gave very similar results: 10  $\mu$ M lactacystin  $\beta$ -lactone (Calbiochem) and 50  $\mu$ M Ac-LLN (Sigma) stabilized the 2XUb-Bla fusion protein and caused the accumulation of high molecular weight ubiquitin conjugates.

5 Proteins destined to be degraded by the proteasome are initially modified by the covalent addition of ubiquitin to lysines within the targeted protein through an isopeptide linkage between the C-terminal residue of ubiquitin and the  $\epsilon$ -amino groups of the substrate protein. The conjugated ubiquitin(s) acts as a high affinity conjugation site for the addition of additional ubiquitin polypeptides in isopeptide  
10 linkage between the C-terminus of the incoming ubiquitin to a lysine residue within the conjugated ubiquitin. When the ubiquitin chains reach a critical size four or more ubiquitin residues long (see Thrower et al., (2000) EMBO J. 19 94-102)), the ubiquitin-protein conjugate is recognized by the proteasome with high affinity, the substrate protein is degraded and the ubiquitin residues are recycled for further rounds  
15 of ubiquitination. To test whether poly-ubiquitination is required for the degradation of 2XUb-Bla, we used a form of ubiquitin where all amines had been reductively methylated, thereby producing a form of ubiquitin that can be conjugated but not extended. When methylated ubiquitin (MeUb) was added to the *in vitro* degradation system at 1 mg/ml, it significantly stabilized 2XUb-Bla and resulted in the appearance  
20 of ladders of labeled species that contain low (1-5 copies) numbers of conjugated ubiquitin polypeptides. (**FIG. 10**) It also inhibited the formation of the very high molecular weight ubiquitin-substrate conjugates observed with the proteasome inhibitors. Collectively, the *in vitro* inhibitor data demonstrate that the multiubiquitin destabilization domain targets degradation of the protein it is fused to in a  
25 proteasome-dependent manner that requires poly-ubiquitination of the substrate for high efficiency recognition/degradation.

Jurkat cells expressing 2XUb-Bla fusion protein were used to test several inhibitors of proteasome function that were active in the *in vitro* system to determine if they were also active within living cells. Cells were treated with various  
30 concentrations of the proteasome inhibitors MG132 or Ac-LLN for 30 minutes at 37°C and then cycloheximide was added to 100  $\mu$ g/ml to initiate a chase period. After 1 hour at 37°C, the cells were cooled to room temperature and then loaded with 1  $\mu$ M

CCF2-AM and  $\beta$ -lactamase activity quantified using a CytoFluor plate reader. The background-subtracted emission values at 460 nm and 530 nm were expressed as a 460/530 ratio and dose-response curves were plotted. **FIG. 11** shows that both MG132 and Ac-LLN exhibited a dose-dependent inhibition of the decay of  $\beta$ -  
5 lactamase activity indicating that they had inhibited the intracellular degradation of the ubiquitin- $\beta$ -lactamase fusion protein.  $IC_{50}$  values calculated from linear regression analysis were found to be 13  $\mu$ M for Ac-LLN and 2.1  $\mu$ M for MG132 and are within the characteristic range for inhibition of substrates degraded by the proteasome (see Lee and Goldberg, (1998) Trends Cell Biol., 8 397-403). These data  
10 demonstrate that the multiubiquitin destabilization domain fused to  $\beta$ -lactamase can serve as a robust cell-based 96-well format screening assay for inhibitors of the proteasome.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described  
15 method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for  
20 carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

## SEQUENCE ID. LISTING

### (1) GENERAL INFORMATION:

5

(iii) NUMBER OF SEQUENCES: 72

### (2) INFORMATION FOR SEQ. ID. NO. 1:

10

#### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

15

(B) TYPE: peptide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

(B) LOCATION: 1....6

25

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.:1:

DSGLDS

### (2) INFORMATION FOR SEQ. ID. NO.: 2:

30

#### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 228 base pairs

35

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

45

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....228

## 5

10

## 15

20

25

30

35

40

45

50

	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
	CTG	CTA	TGT	GGC	GCG	GTA	TTA	TCC	CGT	GTT	GAC	GCC	GGG	CAA	GAG	CAA	CTC
	Leu	Leu	Cys	Gly	Ala	Val	Leu	Ser	Arg	Val	Asp	Ala	Gly	Gln	Glu	Gln	Leu
5																	
		210			220			230			240			250			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
10	GGT	CGC	CGC	ATA	CAC	TAT	TCT	CAG	AAT	GAC	TTG	GTT	GAG	TAC	TCA	CCA	GTC
	Gly	Arg	Arg	Ile	His	Tyr	Ser	Gln	Asn	Asp	Leu	Val	Glu	Tyr	Ser	Pro	Val
		260			270			280			290			300			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
15	ACA	GAA	AAG	CAT	CTT	ACG	GAT	GGC	ATG	ACA	GTA	AGA	GAA	TTA	TGC	AGT	GCT
	Thr	Glu	Lys	His	Leu	Thr	Asp	Gly	Met	Thr	Val	Arg	Glu	Leu	Cys	Ser	Ala
		310			320			330			340			350			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
20	GCC	ATA	ACC	ATG	AGT	GAT	AAC	ACT	GCG	GCC	AAC	TTA	CTT	CTG	ACA	ACG	ATC
	Ala	Ile	Thr	Met	Ser	Asp	Asn	Thr	Ala	Ala	Asn	Leu	Leu	Leu	Thr	Thr	Ile
		360			370			380			390			400			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
25	GGA	GGA	CCG	AAG	GAG	CTA	ACC	GCT	TTT	TTG	CAC	AAC	ATG	GGG	GAT	CAT	GTA
	Gly	Gly	Pro	Lys	Glu	Leu	Thr	Ala	Phe	Leu	His	Asn	Met	Gly	Asp	His	Val
		410			420			430			440			450			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
30	ACT	CGC	CTT	GAT	CGT	TGG	GAA	CCG	GAG	CTG	AAT	GAA	GCC	ATA	CCA	AAC	GAC
	Thr	Arg	Leu	Asp	Arg	Trp	Glu	Pro	Glu	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp
		460			470			480			490			500			510
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
35	GAG	CGT	GAC	ACC	ACG	ATG	CCT	GCA	GCA	ATG	GCA	ACA	ACG	TTG	CGC	AAA	CTA
	Glu	Arg	Asp	Thr	Thr	Met	Pro	Ala	Ala	Met	Ala	Thr	Thr	Leu	Arg	Lys	Leu
		520			530			540			550			560			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
40	TTA	ACT	GGC	GAA	CTA	CTT	ACT	CTA	GCT	TCC	CGG	CAA	CAA	TTA	ATA	GAC	TGG
	Leu	Thr	Gly	Glu	Leu	Leu	Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile	Asp	Trp
		570			580			590			600			610			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
45	ATG	GAG	GCG	GAT	AAA	GTT	GCA	GGA	CCA	CTT	CTG	CGC	TCG	GCC	CTT	CCG	GCT
	Met	Glu	Ala	Asp	Lys	Val	Ala	Gly	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala
		620			630			640			650			660			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
50	GGC	TGG	TTT	ATT	GCT	GAT	AAA	TCT	GGA	GCC	GGT	GAG	CGT	GGG	TCT	CGC	GGT
	Gly	Trp	Phe	Ile	Ala	Asp	Lys	Ser	Gly	Ala	Gly	Glu	Arg	Gly	Ser	Arg	Gly
		670			680			690			700			710			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
55	ATC	ATT	GCA	GCA	CTG	GGG	CCA	GAT	GGT	AAG	CCC	TCC	CGT	ATC	GTA	GTT	ATC
	Ile	Ile	Ala	Ala	Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile	Val	Val	Ile
		720			730			740			750			760			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
60	TAC	ACG	ACG	GGG	AGT	CAG	GCA	ACT	ATG	GAT	GAA	CGA	AAT	AGA	CAG	ATC	GCT
	Tyr	Thr	Thr	Gly	Ser	Gln	Ala	Thr	Met	Asp	Glu	Arg	Asn	Arg	Gln	Ile	Ala
		770			780			790									
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
65	GAG	ATA	GGT	GCC	TCA	CTG	ATT	AAG	CAT	TGG							
	Glu	Ile	Gly	Ala	Ser	Leu	Ile	Lys	His	Trp							



(2) INFORMATION FOR SEQ. ID. NO.:4:

## 5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 858 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20

(B) LOCATION: 1.....858

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO. 4:

25 range 1 to 858

				10			20			30			40			50	
		*		*	*		*	*	*	*	*	*	*	*	*	*	*
30	ATG	AGA	ATT	CAA	CAT	TTC	CGT	GTC	GCC	CTT	ATT	CCC	TTT	TTT	GCG	GCA	TTT
	Met	Arg	Ile	Gln	His	Phe	Arg	Val	Ala	Leu	Ile	Pro	Phe	Phe	Ala	Ala	Phe
				60			70			80			90				100
		*		*	*		*	*	*	*	*	*	*	*	*	*	*
35	TGC	CTT	CCT	GTT	TTT	GGT	CAC	CCA	GAA	ACG	CTG	GTG	AAA	GTA	AAA	GAT	GCT
	Cys	Leu	Pro	Val	Phe	Gly	His	Pro	Glu	Thr	Leu	Val	Lys	Val	Lys	Asp	Ala
				110			120			130			140				150
	*		*	*	*		*	*	*	*	*	*	*	*	*	*	*
40	GAA	GAT	CAG	TTG	GGT	GCA	CGA	GTG	GGT	TAC	ATC	GAA	CTG	GAT	CTC	AAC	AGC
	Glu	Asp	Gln	Leu	Gly	Ala	Arg	Val	Gly	Tyr	Ile	Glu	Leu	Asp	Leu	Asn	Ser
				160			170			180			190				200
	*		*	*	*		*	*	*	*	*	*	*	*	*	*	*
45	GGT	AAG	ATC	CTT	GAG	AGT	TTT	CGC	CCC	GAA	GAA	CGT	TTT	CCA	ATG	ATG	AGC
	Gly	Lys	Ile	Leu	Glu	Ser	Phe	Arg	Pro	Glu	Glu	Arg	Phe	Pro	Met	Met	Ser
				210			220			230			240				250
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
50	ACT	TTT	AAA	GTT	CTG	CTA	TGT	GGC	GCG	GTA	TTA	TCC	CGT	GTT	GAC	GCC	GGG
	Thr	Phe	Lys	Val	Leu	Leu	Cys	Gly	Ala	Val	Leu	Ser	Arg	Val	Asp	Ala	Gly
				260			270			280			290				300
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
55	CAA	GAG	CAA	CTC	GGT	CGC	CGC	ATA	CAC	TAT	TCT	CAG	AAT	GAC	TTG	GTT	GAG
	Gln	Glu	Gln	Leu	Gly	Arg	Arg	Ile	His	Tyr	Ser	Gln	Asn	Asp	Leu	Val	Glu

```

      310      320      330      340      350
      *      *      *      *      *
5  TAC TCA CCA GTC ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA
   Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu

      360      370      380      390      400
      *      *      *      *      *
10 TTA TGC AGT GCT GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT
   Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu

      410      420      430      440      450
      *      *      *      *      *
15 CTG ACA ACG ATC GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG
   Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met

      460      470      480      490      500      510
      *      *      *      *      *      *
20 GGG GAT CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC
   Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala

      520      530      540      550      560
      *      *      *      *      *
25 ATA CCA AAC GAC GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG
   Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr

      570      580      590      600      610
      *      *      *      *      *
30 TTG CGC AAA CTA TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA
   Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln

      620      630      640      650      660
      *      *      *      *      *
35 TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG
   Leu Ile Asp Trp Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser

      670      680      690      700      710
      *      *      *      *      *
40 GCC CTT CCG GCT GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT
   Ala Leu Pro Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg

      720      730      740      750      760
      *      *      *      *      *
45 GGG TCT CGC GGT ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT
   Gly Ser Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg

      770      780      790      800      810
      *      *      *      *      *
50 ATC GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT
   Ile Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn

      820      830      840      850
      *      *      *      *
55 AGA CAG ATC GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
   Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp

```

(2) INFORMATION FOR SEQ. ID. NO.: 5:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 795 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

5

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

10 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....795

15

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 5:

range 1 to 795

20                   10                   20                   30                   40                   50  
          \*           \*           \*           \*           \*           \*           \*  
ATG GGG CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG  
Met Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu  
25                   60                   70                   80                   90                   100  
          \*           \*           \*           \*           \*           \*           \*           \*  
GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT  
Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu  
30                   110                   120                   130                   140                   150  
          \*           \*           \*           \*           \*           \*           \*           \*  
GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT  
Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val  
35                   160                   170                   180                   190                   200  
210  
          \*           \*           \*           \*           \*           \*           \*           \*  
CTG CTA TGT GGC GCG GTA TTA TCC CGT GAT GAC GCC GGG CAA GAG CAA CTC  
Leu Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu  
40                   220                   230                   240                   250                   260  
          \*           \*           \*           \*           \*           \*           \*           \*  
GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC  
Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val  
45                   270                   280                   290                   300                   310  
          \*           \*           \*           \*           \*           \*           \*           \*  
ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT  
Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala  
50                   320                   330                   340                   350                   360  
          \*           \*           \*           \*           \*           \*           \*           \*  
GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC  
Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile  
55                   370                   380                   390                   400                   410  
          \*           \*           \*           \*           \*           \*           \*           \*  
GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA  
Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val  
60

		420		430		440		450		460							
	*	*	*	*	*	*	*	*	*	*	*						
5	ACT	CGC	CTT	GAT	CAT	TGG	GAA	CCG	GAG	CTG	AAT	GAA	GCC	ATA	CCA	AAC	GAC
	Thr	Arg	Leu	Asp	His	Trp	Glu	Pro	Glu	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp
		470		480		490		500		510							
	*	*	*	*	*	*	*	*	*	*	*						
10	GAG	CGT	GAC	ACC	ACG	ATG	CCT	GTA	GCA	ATG	GCA	ACA	ACG	TTG	CGC	AAA	CTA
	Glu	Arg	Asp	Thr	Thr	Met	Pro	Val	Ala	Met	Ala	Thr	Thr	Leu	Arg	Lys	Leu
		520		530		540		550		560							
	*	*	*	*	*	*	*	*	*	*	*						
15	TTA	ACT	GGC	GAA	CTA	CTT	ACT	CTA	GCT	TCC	CGG	CAA	CAA	TTA	ATA	GAC	TGG
	Leu	Thr	Gly	Glu	Leu	Leu	Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile	Asp	Trp
		570		580		590		600		610							
	*	*	*	*	*	*	*	*	*	*	*						
20	ATG	GAG	GCG	GAT	AAA	GTT	GCA	GGA	CCA	CTT	CTG	CGC	TCG	GCC	CTT	CCG	GCT
	Met	Glu	Ala	Asp	Lys	Val	Ala	Gly	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala
		620		630		640		650		660							
	*	*	*	*	*	*	*	*	*	*	*						
25	GGC	TGG	TTT	ATT	GCT	GAT	AAA	TCT	GGA	GCC	GGT	GAG	CGT	GGG	TCT	CGC	GGT
	Gly	Trp	Phe	Ile	Ala	Asp	Lys	Ser	Gly	Ala	Gly	Glu	Arg	Gly	Ser	Arg	Gly
		670		680		690		700		710		720					
	*	*	*	*	*	*	*	*	*	*	*	*					
30	ATC	ATT	GCA	GCA	CTG	GGG	CCA	GAT	GGT	AAG	CCC	TCC	CGT	ATC	GTA	GTT	ATC
	Ile	Ile	Ala	Ala	Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile	Val	Val	Ile
			730		740		750		760		770						
	*	*	*	*	*	*	*	*	*	*	*						
35	TAC	ACG	ACG	GGG	AGT	CAG	GCA	ACT	ATG	GAT	GAA	CGA	AAT	AGA	CAG	ATC	GCT
	Tyr	Thr	Thr	Gly	Ser	Gln	Ala	Thr	Met	Asp	Glu	Arg	Asn	Arg	Gln	Ile	Ala
			780		790												
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
40	GAG	ATA	GGT	GCC	TCA	CTG	ATT	AAG	CAT	TGG							
	Glu	Ile	Gly	Ala	Ser	Leu	Ile	Lys	His	Trp							

## (2) INFORMATION FOR SEQ. ID. NO.: 6:

### 45 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 792 base pairs

(B) TYPE: nucleic acid

50

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

### 55 (ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....792

5 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 6:

range 1 to 792

```
10      10      20      30      40      50
      *      *      *      *      *      *
ATG GAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG GGT
Met Asp Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu Gly

15      60      70      80      90      100
      *      *      *      *      *      *
GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG
Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu Glu

20      110     120     130     140     150
      *      *      *      *      *      *
AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT CTG
Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val Leu

25      160     170     180     190     200
      *      *      *      *      *      *
CTA TGT GGC GCG GTA TTA TCC CGT ATT GAC GCC GGG CAA GAG CAA CTC GGT
Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly

30      210     220     230     240     250
      *      *      *      *      *      *
CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA
Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr

35      260     270     280     290     300
      *      *      *      *      *      *
GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC
Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala

40      310     320     330     340     350
      *      *      *      *      *      *
ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC GGA
Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly

45      360     370     380     390     400
      *      *      *      *      *      *
GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA ACT
Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr

50      410     420     430     440     450
      *      *      *      *      *      *
CGC CTT GAT CAT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG
Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu

55      460     470     480     490     500     510
      *      *      *      *      *      *
CGT GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA
Arg Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu

60      520     530     540     550     560
      *      *      *      *      *      *
ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met
```

```

      570      580      590      600      610
      *      *      *      *      *
GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC
5  Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly

      620      630      640      650      660
      *      *      *      *      *
TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT ATC
10 Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile

      670      680      690      700      710
      *      *      *      *      *
ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC TAC
15 Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr

      720      730      740      750      760
      *      *      *      *      *
ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT GAG
20 Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu

      770      780      790
      *      *      *
ATA GGT GCC TCA CTG ATT AAG CAT TGG
25 Ile Gly Ala Ser Leu Ile Lys His Trp

```

## (2) INFORMATION FOR SEQ. ID. NO.: 7:

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 786 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

### (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....786

### (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 7:

range 1 to 786

```

      10      20      30      40      50
      *      *      *      *      *
ATG AAA GAT GAT TTT GCA AAA CTT GAG GAA CAA TTT GAT GCA AAA CTC GGG
55 Met Lys Asp Asp Phe Ala Lys Leu Glu Glu Gln Phe Asp Ala Lys Leu Gly

      60      70      80      90     100

```

	*	*		*		*	*	*	*	*	*	*	*	*	*	*	*
	ATC	TTT	GCA	TTG	GAT	ACA	GGT	ACA	AAC	CGG	ACG	GTA	GCG	TAT	CGG	CCG	GAT
	Ile	Phe	Ala	Leu	Asp	Thr	Gly	Thr	Asn	Arg	Thr	Val	Ala	Tyr	Arg	Pro	Asp
5			110			120			130			140				150	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GAG	CGT	TTT	GCT	TTT	GCT	TCG	ACG	ATT	AAG	GCT	TTA	ACT	GTA	GGC	GTG	CTT
	Glu	Arg	Phe	Ala	Phe	Ala	Ser	Thr	Ile	Lys	Ala	Leu	Thr	Val	Gly	Val	Leu
10			160			170			180			190				200	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	TTG	CAA	CAG	AAA	TCA	ATA	GAA	GAT	CTG	AAC	CAG	AGA	ATA	ACA	TAT	ACA	CGT
	Leu	Gln	Gln	Lys	Ser	Ile	Glu	Asp	Leu	Asn	Gln	Arg	Ile	Thr	Tyr	Thr	Arg
15			210			220			230			240				250	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GAT	GAT	CTT	GTA	AAC	TAC	AAC	CCG	ATT	ACG	GAA	AAG	CAC	GTT	GAT	ACG	GGA
	Asp	Asp	Leu	Val	Asn	Tyr	Asn	Pro	Ile	Thr	Glu	Lys	His	Val	Asp	Thr	Gly
20			260			270			280			290				300	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	ATG	ACG	CTC	AAA	GAG	CTT	GCG	GAT	GCT	TCG	CTT	CGA	TAT	AGT	GAC	AAT	GCG
	Met	Thr	Leu	Lys	Glu	Leu	Ala	Asp	Ala	Ser	Leu	Arg	Tyr	Ser	Asp	Asn	Ala
25			310			320			330			340				350	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GCA	CAG	AAT	CTC	ATT	CTT	AAA	CAA	ATT	GGC	GGA	CCT	GAA	AGT	TTG	AAA	AAG
	Ala	Gln	Asn	Leu	Ile	Leu	Lys	Gln	Ile	Gly	Gly	Pro	Glu	Ser	Leu	Lys	Lys
30			360			370			380			390				400	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GAA	CTG	AGG	AAG	ATT	GGT	GAT	GAG	GTT	ACA	AAT	CCC	GAA	CGA	TTC	GAA	CCA
	Glu	Leu	Arg	Lys	Ile	Gly	Asp	Glu	Val	Thr	Asn	Pro	Glu	Arg	Phe	Glu	Pro
35			410			420			430			440				450	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GAG	TTA	AAT	GAA	GTG	AAT	CCG	GGT	GAA	ACT	CAG	GAT	ACC	AGT	ACA	GCA	AGA
	Glu	Leu	Asn	Glu	Val	Asn	Pro	Gly	Glu	Thr	Gln	Asp	Thr	Ser	Thr	Ala	Arg
40			460			470			480			490				500	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GCA	CTT	GTC	ACA	AGC	CTT	CGA	GCC	TTT	GCT	CTT	GAA	GAT	AAA	CTT	CCA	AGT
	Ala	Leu	Val	Thr	Ser	Leu	Arg	Ala	Phe	Ala	Leu	Glu	Asp	Lys	Leu	Pro	Ser
45			520			530			540			550				560	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GAA	AAA	CGC	GAG	CTT	TTA	ATC	GAT	TGG	ATG	AAA	CGA	AAT	ACC	ACT	GGA	GAC
	Glu	Lys	Arg	Glu	Leu	Leu	Ile	Asp	Trp	Met	Lys	Arg	Asn	Thr	Thr	Gly	Asp
50			570			580			590			600				610	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GCC	TTA	ATC	CGT	GCC	GGA	GCG	GCA	TCA	TAT	GGA	ACC	CGG	AAT	GAC	ATT	GCC
	Ala	Leu	Ile	Arg	Ala	Gly	Val	Pro	Asp	Gly	Trp	Glu	Val	Ala	Asp	Lys	Thr
55			620			630			640			650				660	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	ATC	ATT	TGG	CCG	CCA	AAA	GGA	GAT	CCT	GTC	GGT	GTG	CCG	GAC	GGT	TGG	GAA
	Gly	Ala	Ala	Ser	Tyr	Lys	Gly	Asp	Pro	Val	Gly	Thr	Arg	Asn	Asp	Ile	Ala
60			670			680			690			700				710	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	GTG	GCT	GAT	AAA	ACT	GTT	CTT	GCA	GTA	TTA	TCC	AGC	AGG	GAT	AAA	AAG	GAC
	Ile	Ile	Trp	Pro	Pro	Val	Leu	Ala	Val	Leu	Ser	Ser	Arg	Asp	Lys	Lys	Asp
65			720			730			740			750				760	

\* \* \* \* \*  
GCC AAG TAT GAT GAT AAA CTT ATT GCA GAG GCA ACA AAG GTG GTA ATG AAA  
Ala Lys Tyr Asp Asp Lys Leu Ile Ala Glu Ala Thr Lys Val Val Met Lys

5           770               780  
          \*               \*               \*               \*  
GCC TTA AAC ATG AAC GGC AAA  
Ala Leu Asn Met Asn Gly Lys

10           (2) INFORMATION FOR SEQ. ID. NO. 8:

(i) SEQUENCE CHARACTERISTICS:

15   (A) LENGTH: 720 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....720

30   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO. 8:

ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC  
35   GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC  
GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC  
40   GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC TTC TCC TAC GGC  
GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC  
AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG  
45   GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC  
CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC  
50   ATC CTG GGG CAC AAC CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC  
ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC  
AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC  
55   CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC  
CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG



CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC  
AAG TAA

5

(2) INFORMATION FOR SEQ. ID. NO.: 9:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 690 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

25 (B) LOCATION: 1....690

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.9:

30 ATG GCT CTT TCA AAC AAG TTT ATC GGA GAT GAC ATG AAA ATG ACC TAC CAT  
ATG GAT GGC TGT GTC AAT GGG CAT TAC TTT ACC GTC AAA GGT GAA GGC AAC  
GGG AAG CCA TAC GAA GGG ACG CAG ACT TCG ACT TTT AAA GTC ACC ATG GCC  
35 AAC GGT GGG CCC CTT GCA TTC TCC TTT GAC ATA CTA TCT ACA GTG TTC AAA  
TAT GGA AAT CGA TGC TTT ACT GCG TAT CCT ACC AGT ATG CCC GAC TAT TTC  
40 AAA CAA GCA TTT CCT GAC GGA ATG TCA TAT GAA AGG ACT TTT ACC TAT GAA  
GAT GGA GGA GTT GCT ACA GCC AGT TGG GAA ATA AGC CTT AAA GGC AAC TGC  
TTT GAG CAC AAA TCC ACG TTT CAT GGA GTG AAC TTT CCT GCT GAT GGA CCT  
45 GTG ATG GCG AAG AAG ACA ACT GGT TGG GAC CCA TCT TTT GAG AAA ATG ACT  
GTC TGC GAT GGA ATA TTG AAG GGT GAT GTC ACC GCG TTC CTC ATG CTG CAA  
50 GGA GGT GGC AAT TAC AGA TGC CAA TTC CAC ACT TCT TAC AAG ACA AAA AAA  
CCG GTG ACG ATG CCA CCA AAC CAT GTG GTG GAA CAT CGC ATT GCG AGG ACC  
GAC CTT GAC AAA GGT GGC AAC AGT GTT CAG CTG ACG GAG CAC GCT GTT GCA  
55 CAT ATA ACC TCT GTT GTC CCT TTC TGA

(2) INFORMATION FOR SEQ. ID. NO.10:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 696 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20 (B) LOCATION: 1....696

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.10:

25 ATG GCT CAG TCA AAG CAC GGT CTA ACA AAA GAA ATG ACA ATG AAA TAC CGT  
ATG GAA GGG TGC GTC GAT GGA CAT AAA TTT GTG ATC ACG GGA GAG GGC ATT  
30 GGA TAT CCG TTC AAA GGG AAA CAG GCT ATT AAT CTG TGT GTG GTC GAA GGT  
GGA CCA TTG CCA TTT GCC GAA GAC ATA TTG TCA GCT GCC TTT AAC TAC GGA  
AAC AGG GTT TTC ACT GAA TAT CCT CAA GAC ATA GTT GAC TAT TTC AAG AAC  
35 TCG TGT CCT GCT GGA TAT ACA TGG GAC AGG TCT TTT CTC TTT GAG GAT GGA  
GCA GTT TGC ATA TGT AAT GCA GAT ATA ACA GTG AGT GTT GAA GAA AAC TGC  
ATG TAT CAT GAG TCC AAA TTT TAT GGA GTG AAT TTT CCT GCT GAT GGA CCT  
40 GTG ATG AAA AAG ATG ACA GAT AAC TGG GAG CCA TCC TGC GAG AAG ATC ATA  
CCA GTA CCT AAG CAG GGG ATA TTG AAA GGG GAT GTC TCC ATG TAC CTC CTT  
45 CTG AAG GAT GGT GGG CGT TTA CGG TGC CAA TTC GAC ACA GTT TAC AAA GCA  
AAG TCT GTG CCA AGA AAG ATG CCG GAC TGG CAC TTC ATC CAG CAT AAG CTC  
ACC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAA TGG CAT CTG ACA GAA  
50 CAT GCT ATT GCA TCC GGA TCT GCA TTG CCC TGA

(2) INFORMATION FOR SEQ. ID. NO.11:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 696 base pairs

(B) TYPE: nucleic acid

5

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

15

(B) LOCATION: 1...696

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.11:

20

ATG GCT CAT TCA AAG CAC GGT CTA AAA GAA GAA ATG ACA ATG AAA TAC CAC

ATG GAA GGG TGC GTC AAC GGA CAT AAA TTT GTG ATC ACG GGC GAA GGC ATT

25

GGA TAT CCG TTC AAA GGG AAA CAG ACT ATT AAT CTG TGT GTG ATC GAA GGG

GGA CCA TTG CCA TTT TCC GAA GAC ATA TTG TCA GCT GGC TTT AAG TAC GGA

GAC AGG ATT TTC ACT GAA TAT CCT CAA GAC ATA GTA GAC TAT TTC AAG AAC

30

TCG TGT CCT GCT GGA TAT ACA TGG GGC AGG TCT TTT CTC TTT GAG GAT GGA

GCA GTC TGC ATA TGC AAT GTA GAT ATA ACA GTG AGT GTC AAA GAA AAC TGC

35

ATT TAT CAT AAG AGC ATA TTT AAT GGA ATG AAT TTT CCT GCT GAT GGA CCT

GTG ATG AAA AAG ATG ACA ACT AAC TGG GAA GCA TCC TGC GAG AAG ATC ATG

CCA GTA CCT AAG CAG GGG ATA CTG AAA GGG GAT GTC TCC ATG TAC CTC CTT

40

CTG AAG GAT GGT GGG CGT TAC CGG TGC CAG TTC GAC ACA GTT TAC AAA GCA

AAG TCT GTG CCA AGT AAG ATG CCG GAG TGG CAC TTC ATC CAG CAT AAG CTC

45

CTC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAG TGG CAG CTG ACA GAG

CAT GCT ATT GCA TTC CCT TCT GCC TTG GCC TGA

50 (2) INFORMATION FOR SEQ. ID. NO.12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 699 base pairs

55

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

10 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1...699

15

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.12:

20

25

30

35

40

45

ATG AGT TGT TCC AAG AGT GTG ATC AAG GAA GAA ATG TTG ATC GAT CTT CAT  
CTG GAA GGA ACG TTC AAT GGG CAC TAC TTT GAA ATA AAA GGC AAA GGA AAA  
GGA CAG CCT AAT GAA GGC ACC AAT ACC GTC ACG CTC GAG GTT ACC AAG GGT  
GGA CCT CTG CCA TTT GGT TGG CAT ATT TTG TGC CCA CAA TTT CAG TAT GGA  
AAC AAG GCA TTT GTC CAC CAC CCT GAC AAC ATA CAT GAT TAT CTA AAG CTG  
TCA TTT CCG GAG GGA TAT ACA TGG GAA CGG TCC ATG CAC TTT GAA GAC GGT  
GGC TTG TGT TGT ATC ACC AAT GAT ATC AGT TTG ACA GGC AAC TGT TTC TAC  
TAC GAC ATC AAG TTC ACT GGC TTG AAC TTT CCT CCA AAT GGA CCC GTT GTG  
CAG AAG AAG ACA ACT GGC TGG GAA CCG AGC ACT GAG CGT TTG TAT CCT CGT  
GAT GGT GTG TTG ATA GGA GAC ATC CAT CAT GCT CTG ACA GTT GAA GGA GGT  
GGT CAT TAC GCA TGT GAC ATT AAA ACT GTT TAC AGG GCC AAG AAG GCC GCC  
TTG AAG ATG CCA GGG TAT CAC TAT GTT GAC ACC AAA CTG GTT ATA TGG AAC  
AAC GAC AAA GAA TTC ATG AAA GTT GAG GAG CAT GAA ATC GCC GTT GCA CGC  
CAC CAT CCG TTC TAT GAG CCA AAG AAG GAT AAG TAA

(2) INFORMATION FOR SEQ. ID. NO.13:

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 678 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

5

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....678

10 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.13:

ATG AGG TCT TCC AAG AAT GTT ATC AAG GAG TTC ATG AGG TTT AAG GTT CGC ATG  
15 GAA GGA ACG GTC AAT GGG CAC GAG TTT GAA ATA GAA GGC GAA GGA GAG GGG AGG  
CCA TAC GAA GGC CAC AAT ACC GTA AAG CTT AAG GTA ACC AAG GGG GGA CCT TTG  
CCA TTT GCT TGG GAT ATT TTG TCA CCA CAA TTT CAG TAT GGA AGC AAG GTA TAT  
20 GTC AAG CAC CCT GCC GAC ATA CCA GAC TAT AAA AAG CTG TCA TTT CCT GAA GGA  
TTT AAA TGG GAA AGG GTC ATG AAC TTT GAA GAC GGT GGC GTC GTT ACT GTA ACC  
25 CAG GAT TCC AGT TTG CAG GAT GGC TGT TTC ATC TAC AAG GTC AAG TTC ATT GGC  
GTG AAC TTT CCT TCC GAT GGA CCT GTT ATG CAA AAG AAG ACA ATG GGC TGG GAA  
GCC AGC ACT GAG CGT TTG TAT CCT CGT GAT GGC GTG TTG AAA GGA GAG ATT CAT  
30 AAG GCT CTG AAG CTG AAA GAC GGT GGT CAT TAC CTA GTT GAA TTC AAA AGT ATT  
TAC ATG GCA AAG AAG CCT GTG CAG CTA CCA GGG TAC TAC TAT GTT GAC TCC AAA  
35 CTG GAT ATA ACA AGC CAC AAC GAA GAC TAT ACA ATC GTT GAG CAG TAT GAA AGA  
ACC GAG GGA CGC CAC CAT CTG TTC CTT TAA

40 (2) INFORMATION FOR SEQ. ID. NO.14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 801 base pairs

45

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

50 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

55

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....801

5 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.14:

ATG AAG TGT AAA TTT GTG TTC TGC CTG TCC TTC TTG GTC CTC GCC ATC ACA  
10 AAC GCG AAC ATT TTT TTG AGA AAC GAG GCT GAC TTA GAA GAG AAG ACA TTG  
AGA ATA CCA AAA GCT CTA ACC ACC ATG GGT GTG ATT AAA CCA GAC ATG AAG  
ATT AAG CTG AAG ATG GAA GGA AAT GTA AAC GGG CAT GCT TTT GTG ATC GAA  
15 GGA GAA GGA GAA GGA AAG CCT TAC GAT GGG ACA CAC ACT TTA AAC CTG GAA  
GTG AAG GAA GGT GCG CCT CTG CCT TTT TCT TAC GAT ATC TTG TCA AAC GCG  
20 TTC CAG TAC GGA AAC AGA GCA TTG ACA AAA TAC CCA GAC GAT ATA GCA GAC  
TAT TTC AAG CAG TCG TTT CCC GAG GGA TAT TCC TGG GAA AGA ACC ATG ACT  
TTT GAA GAC AAA GGC ATT GTC AAA GTG AAA AGT GAC ATA AGC ATG GAG GAA  
25 GAC TCC TTT ATC TAT GAA ATT CGT TTT GAT GGG ATG AAC TTT CCT CCC AAT  
GGT CCG GTT ATG CAG AAA AAA ACT TTG AAG TGG GAA CCA TCC ACT GAG ATT  
30 ATG TAC GTG CGT GAT GGA GTG CTG GTC GGA GAT ATT AGC CAT TCT CTG TTG  
CTG GAG GGA GGT GGC CAT TAC CGA TGT GAC TTC AAA AGT ATT TAC AAA GCA  
AAA AAA GTT GTC AAA TTG CCA GAC TAT CAC TTT GTG GAC CAT CGC ATT GAG  
35 ATC TTG AAC CAT GAC AAG GAT TAC AAC AAA GTA ACG CTG TAT GAG AAT GCA  
GTT GCT CGC TAT TCT TTG CTG CCA AGT CAG GCC TAG

40 (2) INFORMATION FOR SEQ. ID. NO.15:

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 37 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

50 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

55 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.15:

GATCGGTACCACCATGGAGATCTTCGTGAAGACTCTG

(2) INFORMATION FOR SEQ. ID. NO.16:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.16:

20 TGCAGGATCCGTGCATCCCACCTCTGAGACGGAGTACCAG

(2) INFORMATION FOR SEQ. ID. NO.17: (UbiquitinG76V)

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 228 nucleotides

30 (B) TYPE:

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.17:

40 ATG GAG ATC TTC GTG AAG ACT CTG ACT GGT AAG ACC ATC ACC CTC GAA GTG  
GAG CCG AGT GAC ACC ATT GAG AAT GTC AAG GCA AAG ATC CAA GAC AAG GAA  
GGC ATC CCT CCT GAC CAG CAG AGG TTG ATC TTT GCT GGG AAA CAG CTG GAA  
GAT GGA CGC ACC CTG TCT GAC TAC AAC ATC CAG AAA GAG TCC ACC CTG CAC  
CTG GTA CTC CGT CTC AGA GGT GTG

45 (2) INFORMATION FOR SEQ. ID. NO.18: (Ub5' primer)

(A) LENGTH: 32 nucleotides

(B) TYPE: nucleic acid

50

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: oligonucleotide

(B) LOCATION: 1....32

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.:18:  
10

CGAGATCTACCATGGAAATCTTCGTGAAGACT

15

(2) INFORMATION FOR SEQ. ID. NO.19: (Ub3' primer)

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 22 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

30 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....22

35

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 19:

GGATCCGTGGTGCACACCTCTG

40

(2) INFORMATION FOR SEQ. ID. NO. 20 (BLA5)

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid



(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.20:

GATAGGATCCGGGGCGTGGCTGCACCCAGAAACGCTGGTGAAAGTAAAA

10

(2) INFORMATION FOR SEQ. ID. NO.21: (ABSC107)

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

25

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.21:  
GAACTCTAGATTACCAATGCTTAATCAG

30 (2) INFORMATION FOR SEQ. ID. NO.22: (pcDNA3-Bla)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6180 nucleotides

35

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

40 (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

45

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6180

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:22:

gacggatcgggagatctcccgatcccctatggctgactctcagtacaatctgctctgatgccgatagttaagccagtatctg  
5 ctccctgcttgtgtgttgaggctcgtgagtagtgccgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaatagtaataacacggggcattagttcatagcccatatatggagtccgcgttacataacttacggtaaatggc  
ccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgtcccatagtaacgccaatagggaact  
10 ttcattgacgtcaatgggtggactatttacggtaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgcc  
cctattgacgtcaatgacggtaaatggccgctggcattatgccagtagatgaccttgggactttcctacttggcagtag  
atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagctccacccattgacgtcaatgggagtttgggtggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgtcaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
15 gaaccactgcttactggcttatcgaaattaatacgaactactatagggagaccaagcttggtagcagctcggatccggg  
gcgtggctgacccagaaacgctgggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggtgtacatgaact  
ggatctcaacagcggtaagatccttgagagtttgcgccgaagaacgtttccaatgatgagcacttttaagttctgctatgt  
ggcgcgggtattatcccgtattgacggcggaagagcaactcggctgcgcatacactattctcagaatgacttgggtgagt  
actcaccagtcacagaaaagcatcttaccggatggcatgacagtaagagaattatgcagtgtgccataaccatgagtataa  
cactgcggccaacttactctgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgta  
20 actcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaat  
ggcaacaacgttgcgcaactattactggcgaactacttacttagcttcccggcaacaattaatagactggatggaggcg  
gataaagttgcaggaccacttctgcgctcggcccttcggctggctggttattgctgataaatctggagccggtgagcgtg  
ggctcgcgggtatcattgcagcactggggccagatggaagccctccgtagtctgtagtctacacgacggggagtcagg  
caactatggatgaacgaaatagacagatcgtgagataggtgcctcactgattaagcattggtaactagaggccctattct  
25 atagtgtcacctaatactgtagagctcgtgatcagcctcgaactgtgccttctagtggccagcactgtgttggccctcccc  
gtgccttccttgaccctggaagggtgccactcccactgtcctttctaataaaatgaggaaattgcatcgcattgtctgtagtagg  
tgtcattctattctggggggtgggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggg  
gatgcgggtggctctatggttctgaggcggaagaaccagctgggggtctagggggatccccacgcgccctgtagcg  
gcgcattaagcgcggcggtgtggtgttacgcgcagcgtgaccgtacacttggcagcgccttagcggccgctccttc  
30 gctttctcccttcttctcgcacgttcgccggcttccccgtcaagctctaaatcggggcatccctttagggttccgatttagt  
gctttacggcacctcgaacccaaaaaacttgattagggtgatggttcacgtagtggccatcgcctgatagacggttttcg  
ccctttgacgttgaggtccacgttcttaatagtgactctgttccaaactggaacaacactcaacccatctcggctattcttt  
tgattataagggttttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtg  
aatgtgtgtagtgtaggggtgtgaaagtccccaggctccccaggcaggaagatgcaaagcatgcatctcaattagtcagcaacca  
35 agcaaccaggtgtgaaagtccccaggctccccaggcaggaagatgcaaagcatgcatctcaattagtcagcaacca  
tagtccccgccctaactccgccatccccgccctaactccgccaggtccgccattctccgcccatggctgactaattttt  
ttattatgcagaggccgaggccgctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggtttt  
gcaaaaagctccgggagcttgatatccatttcggtatctgatcaagagacaggatgaggatcgtttcgcattgattgaaca  
gatggattgcacgcaggttctccggcgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctg  
40 ctctgatccgccgtgttccggctgtcagcgcagggggcggcggttcttttgcagaccgacctgtccggtgcctgaat  
gaactgcaggacgaggcagcgcggctatcgtggctggccacgacggcggttcttgcgcagctgtgctcagctgtgtca  
ctgaagcgggaaggactggctgctattggcgaaagtccggggcaggtatcctgtcatctcacctgtcctgccgag  
aaagtatccatcatggctgatgcaatcggcggtgtcatcagcttgcctgacctgccattcgaccaccaagcgaaa  
catcgcacgcagcgagcagctactcggatggaagccggtcttctgcatcaggatgatctggacgaagagcatcaggggct  
45 cgcgccagccgaactgttcgccaggtcaaggcgcgcatcccgcggcgaggatctcgtcgtgacctatggcgatgc  
ctgcttggcgaatatcatggtgaaaatggccgcttttctggattcagcactgtggccggctgggtgtggcggaccgctatc  
aggacatagcgttggtaccctgtgatattgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacggat

cgccgctcccgatcgcagcgcacgccttctatgccttctgacgagttcttctgagcgggactctggggttcgaaatgac  
cgaccaagcgacgccaacctgccatcacgagatttcgattccaccgccgcttctatgaaagggtgggcttcggaatcgtt  
ttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgccaccccaactgtttattgcag  
cttataatggttacaaataaagcaatagcatcacaaattcacaaataaagcatttttctactgcattctagttgtggtttgtcaa  
5 actcatcaatgtatcttatcatgtctgtataccgtcgacctctagcttagagcttggcgtaatcatggatcatagctgttctctgtgt  
aaattgttatccgctcacaaatccacacaacatacagaccggaagcataaagtgtaaagcctgggggtgcctaagtgatgag  
ctaactcacattaattgcgttgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcc  
aacgcgcggggagagcggtttgcgtattggcgctcttccgcttctcgtcactgactcgtcgcgtcggctgttcggct  
gcggcgagcgggtatcagctcactcaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatg  
10 tgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgctgtgctggcgtttttccataggctccgccccctga  
cgagcatcacaaaaatcgacgtcaagtcagaggtggcgaaaccgacaggactataaagataaccaggcgtttccccctg  
gaagtcctctcgtgcgtctcctgttccgacctgcccgttaccggataacctgtccgctttctcccttcgggaagcgtggcg  
ctttctcaatgtcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgtt  
cagcccagccgtcgccttatccgtaactatcgtcttgagtccaacccgtaagacacgacttatcgccactggcagca  
15 gccactggtaacaggattagcagagcgaggtatgtagcggtgctacagagttctgaagtgggtgacctactacggctac  
actagaaggacagtatttggtatctgcgctcgtcgaagccagttacctcggaaaaagagttgtagctcttgatccggcaa  
acaaaccaccgctggtagcgggtggtttttgttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatcct  
ttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatttgggtcatgagattcaaaaaaggatct  
tcacctagatccttttaataaaaaatgaagtttaaatcaatctaagtatatatgagtaacttggctgacagttaccaatgctt  
20 aatcagtgaggcacctatctcagcgtatctgtctatttcgttcacatagttgcctgactccccgtcgtgtagataactacgata  
cgggagggccttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaata  
aaccagccagccggaaggccgagcgcagaagtggcctgcaactttatccgctccatccagcttattaattgttgcgg  
gaagctagagtaagtagttcggcagttatagttgcgcaacgttggccattgctacaggcatcgtggtgtcacgctcgtc  
gtttggtatggcttcattcagctccgggtcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggtta  
25 gctccttcggctcctccgatcgtgtcagaagtaagttggccgagtggtatcactcatggttatggcagcactgcataattctct  
tactgtcatgccatccgtaagatgctttctgtgactggtgagtactcaaccaagtcattctgagaatagtgatcgggcgacc  
gagttgctcttccccggcgtaataacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgtt  
cttcggggcgaaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcaccaactgatcttca  
gcacttttactttaccagcgtttctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaaagggaataaggcgac  
30 acggaaatgttgaatactcactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatattg  
aatgtatttagaaaaataaacaataaggggtccgcgcacattccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.23: (pcDNA3-1XUb-Bla)

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6411 nucleotides

40

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

45

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5 (B) LOCATION: 1.....6411

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO: 23:

range 1 to 6411

```
10      *      10      20      30      40      50
      *      *      *      *      *      *      *
gacggatcgggagatctcccgatcccctatggctcgactctcagtacaatctgctctgatgccgcatagt
15 taagccagtatctgctccctgcttgtgtgttggaggctcgctgagtagtgccgcagcaaaaatttaagcta
caacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgtttttgcgctgcttc
gcgatgtacggggcagatatacgcgttgacattgattattgactagttattaatagtaatacaattacgg
ggtcattagttcatagcccatatatggagttccgcgttacataaacttacggtaaatggccgcctggct
gaccgccccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggga
20 ctttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtacatcaagtgtatc
atatgccaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagtgaca
tgaccttatgggactttcctacttggcagtagacatctacgtatttagtcacgtattaccatggtagtc
ggttttggcagtagacatcaatgggcgtggatagcgggttgactcacggggatttcccaagtctccacca
ttgacgtcaatgggagtttgttttggcaccaaaatacaacgggactttccaaaatgtcgtacaactccg
25 cccattgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaa
ctagagaaccactgcttactggcttatcgaaattaatacgaactcactatagggagacccaagcttggt
accaccatggagatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgac
accattgagaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatc
tttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtcacacctg
30 cacctggtagtccgtctcagagggtgtgcaccacggatccggggcggtggctgcaccagaaacgctgggtg
aaagtaaaagatgctgaagatcagttgggtgcacgagtggtttacatcgaactggatctcaacagcgggt
aagatccttgagagtttctgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgt
ggcgcggtattatcccgtattgacgcgggcaagagcaactcggtcgcgcgcatacactattctcagaat
gacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgc
agtgtctgccataaccatgagtataacactgcccgaacttacttctgacaacgatcggaggaccgaag
35 gagctaaccgctttttgacacacatgggggactgtaactcgccttgatcgttgggaaccggagctg
aatgaagccataacaaaacgacgagcgtgcaccacgatgcctgtagcaatggcaacaacgttgcgcaaa
ctattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaa
gttgacaggaccacttctgcgctcggcccttccggctggctgggtttattgctgataaatctggagccggt
gagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagtattc
40 tacacgacggggagtcaggcaactatggatgaacgaatagacagatcgctgagataggtgcctcactg
attaagcattggtaatctagagggccctattctatagtgctcacctaaatgctagagctcgctgacagc
ctcgactgtgccttctagtgtgccagccatctgttgtttgcccctcccccgctgccttcccttgaccctgga
aggtgccactcccactgtcctttcctaataaaaatgaggaaattgcacgcattgtctgagtaggtgtca
ttctattctgggggtgggtgggtgggcagacagcaagggggaggattgggaagacaatagcaggcatgc
45 tggggatgcggtgggtctctatggcttctgagggcgaaagaaccagctggggctctaggggggtatcccc
cgcgcctctagcggcgcattaagcgcggcggtgtgtgtgtgttacgcgcagcgtgacgctacacttgc
cagcgccctagcggccgctcctttcgtcttcttcccttcccttctcgcacagcttcgcggctttccccg
tcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgacccccaaaaa
acttgattaggggtgatgggtcacgtagtgggccatcgccctgatagacgggtttttcgccttttgacgtt
50 ggagtcacaggttctttaatagtggaactctgttccaaactggaacaacactcaaccctatctcggtcta
ttcttttgatttataagggattttggggatttcggcctatttggttaaaaaatgagctgatttaacaaaa
atthaacgcgaatttaattctgtggaatgtgtgtcagttaggtgtggaaagtccccaggtccccaggg
aggcagaagtagtcaaaagcatgcacatctcaattagtcagcaaccatagtcggcccccctaactccgcc
55 catcccgccccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttta
tgacagaggccgaggccgctctgcctctgagctattccagaagtagtgaggaggcttttttgaggccct
aggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgagg
atcgtttcgcagtgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggtctatt
60 cggcttagactgggcacaaacagacaatcggctgcctctgatgcgcgctgttccggctgtcagcgcaggg
cgcccggttctttttgtcaagaccgacctgtccgggtgcctgaatgaactgcaggacgaggcagcgcg
gctatcgtggctggccacgacgggcttcccttgccgagctgtgctcgacgttgtcactgaagcgggaag
ggactggctgctattgggcgaagtgcgggggcaggatctcctgtcatctcaccttgcctcctgccgagaa
agtatccatcatggctgatgcaatgcggggcgtgcatacgttgatccggctacctgccattcgacca
```

ccaagcgaacatcgcatcgagcgcagcactcgatggaagccggtcttgcgatcaggatgatct  
 ggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcagccgacgg  
 cgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggcgcgttttc  
 tggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggtacccgtga  
 5 tattgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacggtatcgccgctcccg  
 ttcgcagcgcacgccttctatcgcccttcttgacgagttcttctgagcgggactctggggttcgaaatg  
 accgaccaagcgcagcgcgaacctgccatcacgagatttcgattccaccgcccgccttctatgaaaggttg  
 ggcttcggaatcggttttcgggacgcgggctggatgatcctccagcgcggggatctcatgctggagttc  
 ttcgcccaccccaacttggtttattgcagcttataatggttacaaaataaagcaatagcatcacaatttc  
 10 acaataaagcatttttttcaactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcat  
 gtctgtataccgtcgacctctagctagagcttggcgtaatcatgggtcatagctgtttcctgtgtgaaat  
 tgttatccgctcacaattccacacaacatacagagccggaagcataaagtgtaaagcctggggtgcctaa  
 tgagtgcgtaactcacattaattgcgttgccgctcactgcccgcctttccagtcgggaaacctgtcgtgc  
 cagctgcattaatgaatcggcgaacgcgcggggagagggcggtttgcgtattgggcgctcttccgcttcc  
 15 tcgctcactgactcgtcgcgtcggctcgttcggctgcggcgagcggatcagctcactcaaggcggtga  
 atacggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggcc  
 aggaaccgtataaaaaggccgcttgctggcggtttttccataggctccgccccctgacgagcatcaca  
 aatcgacgctcaagtcagaggtggcgaaacccgcagaggactataaagataccaggcggtttcccccctgga  
 agctccctcgtgcgctctcctgttccgacctgcccgttacccgataacctgtccgcctttctcccttcg  
 20 ggaagcgtggcgctttctcaatgctcacgctgtaggatctcagttcgggtgtaggctcgttcgctccaag  
 ctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgccttatccggtaactcgtcttgag  
 tccaacccggtgaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgagg  
 tatgtaggcggtgtctacagagttcttgaagtgggtggcctaactacgggtacactagaaggacagtattt  
 ggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaa  
 25 accaccgctggtagcgggtgggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaa  
 gaagatcctttgatcttttctacgggggtctgacgctcagtggaacgaaaactcacgttaagggattttg  
 gtcatgagattatcaaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaatcaatc  
 taaagtataatgagtaaaacttggctgcagagttaccaatgcttaatcagtgaggcacctatctcagcg  
 atctgtctatttctgttcacatagttgcctgactccccgctcgtgtagataactacgatacgggagggc  
 30 ttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccgggtccagatttatcagca  
 ataaaccagccagccggaagggccgagcgcagaagtggctcctgcaactttatccgcctccatccagctt  
 attaatgttgccgggaagctagagtaagtagttcgccagttaatagtttgccgaacggtgttgccatt  
 gctacaggcatcgtggtgtcacgctcgtcgtttggtagtggcttcattcagctccggttcccaacgatca  
 aggcgagttacatgatcccccatgttgtgcaaaaaagcgggttagctccttcgggtcctccgatcgttgct  
 35 agaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataattctcttactgtcatg  
 ccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtagcgg  
 cgaccgagttgctcttgcccggtcgaatacgggataataccgcgccacatagcagaactttaaaagt  
 ctcatcattggaaaacggttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcg  
 atgtaaccactcgtgcaccaactgatcttcagcatcttttactttcaccagcggttctggggtgagca  
 40 aaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc  
 ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgt  
 atttagaaaaataaacaataaggggttccgcgcacatttccccgaaaagtgccacctgacgtc

45 (2) INFORMATION FOR SEQ. ID. NO.24: (pcDNA3-2XUb-Bla)

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 6678 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6678

5

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 24:

range 1 to 6678

10

gacggatcgggagatctcccgatcccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctgctgagtagtgcgcgagcaaaatctaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
15 ctagtattaatagtaataattacggggctcattagttcatagcccataatggagttccgcgttacataaacttacggtaaatggc  
ccgctggtgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacggcc  
cctattgacgtcaatgacggtaaatggccgctggcattatgccagtagacattatgggactttcacttggcagtag  
atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
20 ggatttcaaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggttaggcgtgacgggtgggaggtctatataagcagagctctctggctaactaga  
gaacccactgcttactggcttaccgaaattaatcagctactatagggagacccaagcttgatatagaattcctgcagccccg  
ggggatctaccatggaatcttctgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggtgatcttctgggaaacagctg  
25 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggtgatcttctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatccggggcgtggctgcacccagaaacgctgggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggtg  
30 gttacatcgaactggatctcaacagcggtaagatccttgagagttttgcggcgaagaacgtttccaatgatgagcacttta  
aagtctgctatgtggcgcggtattatcccgtattgacgcgggcaagagcaactcggctgccgcatacactattctcagaat  
gacttgggtgagtagtaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccataa  
ccatgagtataactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacat  
gggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagccatacacaacgacgagcgtgacaccacg  
35 atgcctgtagcaatggcaacaacgttgcgcaaaactattaactggcgaactacttacttagcttcccggcaacaattaataga  
ctggatggaggcggataaagtgcaggaccacttctgcgctcggccctccggctggctgtgttattgctgataaatctgga  
gccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggttaagccctcccgatcgtagtattctacacg  
acggggagtcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctcactgattaagcattggtaatct  
agagggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctcactgtgccttctagtggcagccatctgt  
40 tgtttgcccctccccgtgccttcttgaccctggaagggtgccactcccactgtcctttcctaataaaatagggaattgcatc  
gcattgtctgagtaggtgtcattctattctgggggggtgggtggggcaggacagcaagggggaggattgggaagacaata  
gcaggcatgctggggatgcgggtgggtctatggcttctgaggcggaaagaaccagctggggctctaggggggtatcccca  
cgcgccctgtagcggcgattaagcgcggcgggtgtggtgttacgcgcagcgtgaccgtacacttgccagcggcccta  
gcgccgctcctttcgtttcttcccttcttctgccacgttcgccggcttccccgtcaagctctaaatcggggcatcccttt  
45 aggggtccgatttagtgctttacggcacctcgacccccaaaaaacttgattagggtgatggttcacgtagtggccatcgccct  
gatagacgggttttgccttttgacgttggagtcacgttcttaatagtggactctgttccaaactggaacaacactcaacc  
tatctcggctctattctttgattataagggattttggggatttcggcctatttggttaaaaaatgagctgatttaacaaaaattaac  
gcgaattaattctgtggaatgtgtgtcagttagggtgtggaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcactc  
atgcactcaattagtcagcaaccaggtgtggaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcactc

aattagtcagcaaccatagtcgcccccctaactccgcccatcccgcccctaactccgccagttccgccattctccgccc  
atggctgactaattttttatgtatgcagagccgagggccgctctgctctgagctattccagaagtagtgaggaggcttttt  
ggaggccataggcttttgcaaaaagctccgggagcttgatataccatttccggtatgatcaagagacaggatgaggatcggt  
tcgcatgattgaacaagatggattgcacgcagggttctccggccgcttgggtggagaggctattcggtatgactgggcaca  
5 acagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcagggggcggcggttcttttgcaagaccgacct  
gtccgggtccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggggcgttcttgcgcagct  
gtgctgcagctgtcactgaagcgggaagggactggctgctattggcggaagtggcggggcaggatcctctgcatctcac  
cttgcctcctgccgagaaagtatccatcatggctgatgcaatggcgggcggtgcatacgcttgatccggctacctgccattcg  
accaccaagcgaaacatcgcacgagcagctactcggtatggaagccggcttgcgacgagatgatctggacga  
10 agagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcacgcccagcggcgaggatctcgtcgt  
gacccatggcgatgcctgcttgcgaatatcatggtggaaaatggccgctttctggattatcgactgtggccggctgggt  
gtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcgggcgaatgggtgaccgcttc  
ctcgtgctttacgggtatcgccgtccccgattcgcagcgcacgccttctatcgcttcttgacgagttcttgcagcgggactct  
gggggtcgaatgaccgaccaagcgacgcccacctgccatcacgagatttcgattccaccggcgcttctatgaaagggt  
15 gggcttcggaatcgttttccgggacggcggtggatgactcctccagcgcggggatctcatgctggagtcttccgccaccc  
caactgtttattgcagcttataatggttacaaaataagcaatagcatcacaaatttcacaaataaagcatttttctactgcattct  
agttgtggtttgccaactcatcaatgtatcttatcatgctgtataccgtcgacctctagctagagcttggcgtaatcatggtc  
atagctgtttctgtgtgaaattgttatccgctcacaaattccacacaacatacagagccgggaagcataaagtgtaaagcctggg  
gtgcctaatagtgagtaactcacattaattgcgttgcgctcactccccgcttccagtcgggaaacctgtcgtgccagctg  
20 catlaatgaatggccaacgcgcggggagaggcggttgcgtattggcgcttctccgcttctcgtcactgactcgtcgc  
gctcggctgttcggctcggcgagcggtatcagctcactcaaaggcggtatacgggtatccacagaatcaggggataac  
gcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcggttttccata  
ggctccgccccctgacgagcatcacaaaaatcgcgctcaagtcagaggtggcgaaaccggacaggactataaagata  
ccaggcggttccccctggaagctccctcgtgcgctcctgttccgacctgccgcttaccggataacctgtccgctttctccc  
25 ttccgggaagcgtggcgctttctcaatgctcagctgtaggtatctcagttcggtgtaggtcgttcgctccaaagctgggctgtg  
gcacgaacccccgttcagcccgaccgctgcgcttatccggtaactatcgtcttgagttcaaccggtaagacacgactt  
atcgccactggcagcagccactggtaacaggattagcagagcgagggtatgtaggcggtgctacagagttctgaagtgg  
ggcctaactacggctacactagaaggacagtatttggtatctgcgctcgtctgaagccagttaccttcggaaaaagagtgg  
agctcttgatccggcaacaaaccaccgctggtagcggtgttttttttgaagcagcagattacgcgcagaaaaaag  
30 gatctcaagaagatccttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggtttgtgcatga  
gattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagtttaaatcaatctaaagtatatatgagtaacttgg  
ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgtatctgtctatttcgttcacatagttgctgactccccgtc  
gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc  
tccagattatcagcaataaaccagccagccggaagggccgagcgcagaagtggctcctgaactttatccgctccatcca  
35 gtctattaattgttggcgggaagctagagtaagtagttcgccagttaatagtttgcgaacgttgttgcattgctacaggcatc  
gtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttg  
tgcaaaaaagcggttagctccttcggtcctccgacgttgcagaagtaagttggcgcagtggtatcactcatggttatggca  
gcactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaata  
gtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggataatccgcgccacatagcagaactttaaaagtgt  
40 catcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
caccctaactgatcttcagcatcttttaccatccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa  
gggaataagggcgacacggaaatgttgaatactcactcttcttttcaatattattgaagcattatcagggttattgtctcat  
gagcggatacatattgaatgtatttagaaaaataaacaataaggggtccgcgcacattccccgaaaagtgccacctgac  
gtc  
45

(2) INFORMATION FOR SEQ. ID. NO.: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6981 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6981

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 25:

range 1 to 6981

gacggatcgggagatctcccgatcccctatggcgcactctcagtacaatctgctctgatgccgcagtaagccagtatctg  
ctccctgcttgtgtgttggaggctgcgtgagtagtgccgcagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatafacgcgttgacattgattattga  
ctagttattaatagtaataacacggggcattagttcatagcccatatatggagtccgcgttacataacttacggtaaatggc  
ccgctggtgacgcgcccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactatttaccggtaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
atctacgtatttagtcacgctattaccatgggtgatgcgggttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaaccactgcttactggcttatcgaaattaatacactcactataggagaccaagcttgatcgaattctgcagcccg  
ggggatctaccatggaatcttcgtgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
cacggtatctaccatggaatcttcgtgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
cacggtatctaccatggaatcttcgtgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
cacggtatccggggcgtggctgcacccagaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggtg  
gttacatcgaactggatctcaacagcggtaagatccttgagagtttgcgccgaagaacgtttccaatgatgacactttta  
aagttctgctatgtggcgcgggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaat  
gacttgggtgagtagtaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccataa  
ccatgagtataacactgcggccaacttactctgacaacgatcgaggaccgaaggagctaaccgctttttgcacaacat



gggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacg  
 atgcctgtagcaatggcaacaacgttgcgcaaaactattaactggcgaactacttactctagcttcccggcaacaattaataga  
 ctggatggagggcgataaagtgcaggaccacttctgcgctcggccctccggctggctggttattgctgataaatctgga  
 gccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggaagccctcccgatcgtagtattctacacg  
 5 acgggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggctgcctcactgattaagcattggaatct  
 agagggccctattctatagtgacacctaagtctagagctcgtgatcagcctcgaactgtgccttctagtggcagccatctgt  
 tglttgccccctccccgtgccttcttgacctggaagtgccactccactgtcctttcttaataaaatgaggaaattgcatc  
 gcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaata  
 gcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccca  
 10 cgcgccctgtagcggcgcaataagcgcggcggtgtgggtgttacgcgcagcgtgaccgctacacttgccagcgccta  
 gcgcccgtctcttctgcttcttcccttcttctcgcacgttcgccggcttccccgtcaagctctaaatcggggcgcctctt  
 aggggtccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtggccatcgccct  
 gatagacggttttgccttggacgttgagtcacgttcttaatagtggactctgttccaaactggaacaacactcaaccc  
 tatctcggctctattctttgattataagggattttggggatttcggcctattggttaaaaaatgagctgatttaaaaaaatttaac  
 15 gcgaattaattctgtggaatgtgtgtagttagggtgtggaaagtcggcaggtccccaggcaggcagaagtatgcaaagc  
 atgcatctcaattagtcagcaaccagggtgtggaaagtcggcaggtccccagcaggcagaagtatgcaaagcatgcatctc  
 aattagtcagcaaccatagtcggcccttaactcggccatcccgcccttaactcggccagttccgcccatttccgcccc  
 atggctgactaattttttattatgcagaggccgagggcgctctgcctctgagctattccagaagtagtgaggaggcttttt  
 ggaggcctaggcttttcaaaaagctcccgggagcttgtatatccatttccggtctgatcaagagacaggatgaggatcgtt  
 20 tcgcatgattgaacaagatggattgcacgcaggttctccggcgcttgggtggagaggctattcggctatgactgggcaca  
 acagacaatcggctgctctgatccgctgttccggctgtcagcgaggggcgccccggttcttttgcagaccgacct  
 gtccgggtgccctgaatgaactgcaggacgaggcagcgcggtctatcgtggtggccacgacgggcgttcttgcgcagct  
 gtgctcagctgtcactgaagcgggaaggactggctgctattggcggaagtccggggcaggatcctctgtcatctcac  
 cttgctcctgccgagaagatccatcatggctgatgcaatgcggcggtgcatacgcttgatccggctacctgcccattcg  
 25 accaccaagcgaaacatcgcatcgagcgagcacgtactggatggaagccggcttctgctgatcaggatgatctggacga  
 agagcatcaggggctcgcgccagccgaactgttcgccagggtcaaggcgcgcagcccacggcgaggatctcgtcgt  
 gacctatggcgatgcctgcttggcgaatatcatggtggaaatggccgcttcttggattcatcgactgtggccggctgggt  
 gtggcgaggaccgtatcaggacatagcgttggctaccgtgatattgctgaagagcttggcggcgaatgggtgaccgcttc  
 ctgctgtttacgggtatcgccgctccgattcgcagcgcatcgcttctatcgcttcttgacgagttcttgcagcgggactct  
 30 ggggttcgaaatgaccgaccaagcgacgccaacctgccatcacgagatttcgattccaccgccgcttctatgaaagggt  
 gggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttctcggccaccc  
 caactgtttattgcagcttataatggttacaaataaagcaatagcatcaaaattcacaaataaagcatttttactgcattct  
 agttgtggtttgccaactcatcaatgtatcttatcatgtctgtataccgctgacctctagctagagcttggcgtaatcatggtc  
 atagctgtttctgtgtgaaattgtatccgctcacaattccacacaatacagagccggaagcataaagtgtaaagcctggg  
 35 gtgcctaagttagtgagctaactcacattaattgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctg  
 cattaatgaatcgccaacgcgcggggagagggcggtttgcgtattggcgctcttccgcttctcgtcactgactgctgc  
 gctcggctgttcggctcggcgagcgggtatcagctcactcaaaaggcggaatacgggtatccacagaatcaggggataac  
 gcaggaaagaacatgtgagcaaaaggccagcaaaaggccagggaaccgtaaaaaggccgcgttgcgtggcgttttccata  
 ggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccgacaggactataaagata  
 40 ccaggcggtttccccctggaagctccctcgtgcgctctcctgttccgacctgccgcttaccggataacctgtccgctttctccc  
 ttccgggaagcgtggcgcttttcaatgctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgt  
 gcacgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttgagtcgaacccggtaagacacgactt  
 atcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgggtgtacagagttcttgaagtgggt  
 ggcttaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggt  
 45 agctcttgatccggcaacaaaccaccgctggtagcgggtgtttttgtttgcaagcagcagattacgcgcagaaaaaaag  
 gatctcaagaagatcctttgatctttctacgggggtctgacgctcagtggaacgaaaactcacgttaagggaattttggtcatga  
 gattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatcaagatatatgagtaacttgggt

ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacccatagttgcctgactccccgtc  
 gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc  
 tccagattatcagcaataaaccagccagccggaaggccgagcgcagaagtggctctgcaactttatccgctccatcca  
 gtctattaattgttgccgggaagctagagtaagtagttccagttaatagtttgcgcaacgttggtgccattgctacaggcatc  
 5 gtggtgtcacgctcgtcgtttggtatggcttcattcagctccgggtcccaacgatcaaggcgagttacatgatccccatgttg  
 tgcaaaaaagcggttagctccttcggctcctccgatcgttgcagaaagtaagttggccgcagtggtatcactcatggttatggca  
 gcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtagtactcaaccaagtcattctgagaata  
 gtgtatgcggcgaccgagtgctcttgcggcggtcaatacgggataataccgcgccacatagcagaactttaaaagtgtc  
 catcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
 10 cacccaactgatcttcagcatcttttaccagcgtttctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaaa  
 gggaataaggggcgacacggaaatgttgaatactcatacttctcttttcaatattattgaagcattatcagggtattgtctcat  
 gagcggatacatatttgaaatgtatttagaaaaataaacaataagggggtccgcgcacattccccgaaaagtgccacctgac  
 gtc

15 (2) INFORMATION FOR SEQ. ID. NO.: 26:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 7164 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

30 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7164

35 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 26:

range 1 to 7164

gacggatcgggagatctccgatccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
 40 ctccctgcttgtgtgttgaggtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
 gcatgaagaatctgcttagggtaggcgttttcgctgttcgcgatgtacgggccagatatacgcgttgacattgattattga  
 ctagttattaatagtaaatcaattacggggctcattagttcatagccatataatggagttccgcgttacataaactacggtaaatggc  
 ccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
 ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtacatcaagtgtatcatatgccaaagtacggcc  
 45 cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtacatgaccttatgggactttcctacttggcagtac  
 atctacgtattagtcacgtattaccatgggtgatcggttttggcagtacatcaatgggcgtggatagcgggttgactcacgg  
 ggattccaagtctccacccattgacgtcaatgggagtttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga

gaaccactgcttactggcttactgaaattaatacgaactactatagggagaccaagcttgatatgaattcctgcagccc  
 ggggatctaccatggaaatcttctgaagactctgactggaagaccatcactctgaagtggagccgagtacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
 5 cacggatctaccatggaaatcttctgaagactctgactggaagaccatcactctcgaagtggagccgagtacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatcttctgaagactctgactggaagaccatcactctcgaagtggagccgagtacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctg  
 10 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatcttctgaagactctgactggaagaccatcactctcgaagtggagccgagtacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatcttctgaagactctgactggaagaccatcactctcgaagtggagccgagtacaccattg  
 15 gttacatcgaactggatctcaacagcggtgaagatccttgagagtcttgcggcgaagaacgtttccaatgatgagcacttta  
 aagtctgctatgtggcgcggtattatcccgtattgacggcgggcaagagcaactcggcgcgcatacactattctcagaat  
 gacttgggtgagtactaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccataa  
 ccatgagtataactgcggccaacttactctgacaacgatcgaggaccgaaggagctaaccgctttttgcacaacat  
 gggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacg  
 20 atgcctgtagcaatggcaacaacgttgcgcaactattaactggcgaactacttacttagcttcccggcaacaattaataga  
 ctggatggaggggataaagtgacaggaccacttctgcgtcggccctccggctggctgtttattgctgataaatctgga  
 gccggtgagcgtgggtctgcggatcattgcagcactggggccagatggtaagccctccgctatcgtagtattctacacg  
 acggggagtcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctcactgattaagcattgtaatct  
 agagggccctattctatagtgtcacctaaatgtagagctcgtgacagcctcactgtgcttcttagttgccagccatctgt  
 25 tgtttgccccctccccgtgccttcttgacctggaggtgccactcccactgtccttcttaataaaataggaaattgcac  
 gcattgtctgagttaggtgtcattctattctgggggggtgggtggggcaggacagcaagggggaggattgggaagacaata  
 gcaggcatgctggggatgcgggtgggtctatggcttctgaggcggaaagaaccagctggggcttaggggggtatcccca  
 cgcgccctgtagcggcgcaataagcgcggcgggtgtggtgttacgcgcagcgtgaccgctacacttggcagcgcctta  
 gcgcccgtccttctgcttcttcccttcttctcgcacgttcgcccgttccccgtcaagctctaatacggggcatccctt  
 30 aggggtccgatttagtgccttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggcatcgcct  
 gatagacgggttttgccttggacgttggagtcacgttcttaataagtgactctgttccaaactggaacaacactcaacc  
 tatctcggcttattctttgattataagggatttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaat  
 gcgaattaattctgtggaatgtgtgacgttaggtgtggaagtccccaggctccccaggcaggcagaagtatgcaagc  
 atgcatctcaattagtcagcaaccaggtgtggaagtccccaggctccccaggcaggcagaagtatgcaagcatgcatctc  
 35 aattagtcagcaaccatagtcggcccttaactcggccatcccgcccttaactcggccaggtccgcccattctccgccc  
 atggctgactaatttttttattatgcagaggccgaggccgcctctgcctctgagctattccagaagtatgaggaggcttttt  
 ggaggccataggcttttgcaaaaagctcccgggagcttgcctatccatttccgatctgatcaagagacaggatgaggatcgtt  
 tcgcatgattgaacaagatggattgcacgcagggtctccggcgttgggtggagaggctattcggctatgactgggcaca  
 acagacaatcggctgctctgatccgccgtgtccggctgtcagcgcaggggcggcgggttcttttgcgaagaccgacct  
 40 gtccgggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggcggttccttgcgcagct  
 gtgctcgcaggtgtcactgaagcgggaaggactggctgctattgggcgaagtggcggggcaggatctcctgtcatctcac  
 ctgtcctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgttgatccggctacctgccattcg  
 accaccaagcgaacatcgcacgcagcagcgtactcggatggaagccggtctgtcgcacaggtgatctggacga  
 agagcatcaggggctcgcggcagccgaactgttcggcaggctcaaggcgcgcagcggcgaggatctcgtcgt  
 45 gacctatggcgatgcctgcttcccgaatatcatggtggaaaatggcgccttctggattcatcactgtggcgggtgggt  
 gtggcggaccgtatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttc  
 ctcgtgctttacggtatcggcgtcccgttcgcagcgcacgccttctatcgccttcttgacgagttctctgagcgggactc

5 ggggttcgaaatgaccgaccaagcgacgcccacctgccatcacgagatttcgattccaccgccgccttctatgaaaggtt  
 gggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgccaccc  
 caacttgttattgcagcttataatggttacaaataagcaatagcatcacaaattcacaaataaagcattttttactgcattct  
 agttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtc  
 10 atagctgtttcctgtgtgaaattgtatccgctcacaaattccacacaacatacagaccggaagcataaagtgtaaagcctggg  
 gtgcctaagtgtgagtaactcacattaattgcgttgctcactgccgctttccagtcgggaaacctgtcgtgccagctg  
 cattaatgaatcgccaacgcgcggggagagggcggttgcgtattggcgctcttccgcttctcgtcactgactcgtcgc  
 gctcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcggtatacggttatccacagaatcaggggataac  
 gcaggaaaagaacatgtgagcaaaaaggccagcaaaaaggccaggaaccgtaaaaaggccgcgttgctggcgttttccata  
 15 ggctccgccccctgacgagcatcacaaaaatcgactcaagtcagaggtggcgaaacccgacaggactataaagata  
 ccaggcggtttccccctggaagtcctctgctgcgtctctgttccgacctgccgcttaccggataacctgtccgccttttccc  
 ttccgggaagcgtggcgcttttcaatgctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtg  
 gcacgaacccccgttcagcccagccgctgcgccttatccgtaactatcgtcttgagccaacccggaagacacgactt  
 atcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgggtgtacagagttcttgaagtgg  
 20 ggcctaactacggctacactagaaggacagtatttggtatctgcgctctgtaagccagttaccttcgaaaaagagttggt  
 agctcttgatccggcaacaaaccaccgctggtagcgggtggtttttgttgaagcagcagattacgcgcagaaaaaag  
 gatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgtaagggttttggcatga  
 gattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttgg  
 ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacatagttgcctgactccccgtc  
 25 gtgtagataactacgatacgggaggggttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggc  
 tccagatttatcagcaataaaccagccagccggaaggggccgagcgcagaagtggctctgcaactttatccgctccatcca  
 gtctattaattgttccgggaagctagagtaagtagttcgccagttaatagtttgcgaacgttggtccattgctacaggcatc  
 gtggtgtcacgctcgtcgttggtaggttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttg  
 tgcaaaaaagcggtagctccttcggctcctccgatcgttgcagaagtaagttggccgcagtgttatcactcatggttatggca  
 30 gactgcataattctctactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaata  
 gtgtatgcggcgaccgagttgctcttggccggcgtaatacgggataataccgcgccacatagcagaactttaaaagtgt  
 catcattggaacggttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
 cacccaactgatcttcagcatcttttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa  
 ggggaataaggcgacacggaaatgttgaatactcactcttcttttcaatattattgaagcatttatcagggttattgtctcat  
 35 gagcggatacatatttgatgtatttagaaaaataaacaataagggggtccgcgcacatttccccgaaaagtgccacctgac  
 gtc

(2) INFORMATION FOR SEQ. ID. NO.27: (pcDNA3-Ub-Met-Bla)

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6411 base pairs

40

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5

(B) LOCATION: 1.....6411

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:27:

10 gacggatcgggagatctcccgatcccctatggcgcactctcagtacaatctgctctgatgccgcagtagttaagccagtatctg  
ctccctgcttgtgtgttggaggctcgcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaataagtaatacaattacggggctcattagttcatagcccataatggagttccgcgttacataacttacggtaaatggc  
ccgcttggctgaccgccccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
15 ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagcatgacctatgggactttcctacttggcagtag  
atctacgtatttagtcatcgcattaccatgggtgatgcgggttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgacgggtgggaggtctatataagcagagctctctggctaactaga  
20 gaaccactgcttactggcttatcgaaattaatacactcactatagggagaccaagcttggtagccaccatggagatcttcg  
tgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgaccaa  
gacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctggaagatggacgcacctgtctga  
ctacaacatccagaaagagtgccacctgcacctggtactccgtctcagaggtgggatgcacggatccggggcgtggctgc  
accagaaacgctgggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatctcaaca  
25 gcggtaagatccttgagagtttgcggcgaagaacgtttccaatgatgagcacttttaaagtctgtatgtggcgcgggtatt  
atcccgattgacggcgggcaagagcaactcggcgcgcacatacactattctcagaatgacttgggtgagtactcaccagtc  
acagaaaagcatcttaccggatggcatgacagtaagagaattatgagtgctgccataaccatgagtataactgcggcc  
aacttacttctgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgtaactgccttg  
atcgttgggaaccggagctgaatgaagccatacacaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaac  
30 gttgcgcaactattaactggcgaaactacttactctagcttccggcacaataatagactggatggaggcggataaagtgtg  
caggaccacttctgcgctcggcccttccggctggctgtttattgctgataaatctggagccggtagcgtgggtctcgcgg  
tatcattgcagcactggggccagatggtgaagccctcccgtatcgtatgtatctacacgacggggagtcaggcaactatggat  
gaacgaaatagacagatcgtgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtacc  
taaagtctagagctcgtgatcagctcgtctgtgccttctagtgtccagccatctgttgtttgcccctccccgtgccttctt  
35 gaccttggaaaggtgccactcccactgtccttcttaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctatt  
ctgggggtgtgggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgg  
gctctatggcttctgagcggaaagaaccagctggggctctaggggggtatccccacgcgcctgtagcggcgcaataagc  
gcggcgggtgtgtgtgttacgcgcagcgtgaccgtacacttgcagcgcctagcggcgctccttctgcttcttccctt  
ccttctcggcacgttcgccggcttccccgtcaagctctaaatcggggcatcccttaggggtccgatttagtgccttacggca  
40 cctcgacccccaaaaaacttgattaggtgatggttcacgtatgggccatcgcctgatagacgggttttcgcccttgacgtt  
ggagtcacagttcttaataagtgactcttgtccaaactggaacaacactcaaccctatctcggctattcttttgattataagg  
gattttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgca  
gttaggggtgtggaaagtcggcaggtccccaggcaggcagaagtagcaaaagcatgcatctcaattagtcagcaaccagg  
tgtggaaagtcggcaggtccccagcaggcagaagtagcaaaagcatgcatctcaattagtcagcaaccatagtcggcc  
45 cctaactccggccatcccgccctaactcggccaggttcggccattctcggcccatggctgactaattttttatattatgca  
gaggccgaggccgctctgcctctgagctattccagaagtagtgaggaggctttttggaggccctaggcttttgcaaaaagc  
tccggggagcttgtatatccatttccgatctgatcaagagacaggatgaggatcgttcgcatgattgaacaagatggattgc

acgcagggttccggccgcttgggtggagaggctattcggctatgactgggcacacagacaatcggtgctctgatgcc  
 gccgtgtccggctgtcagcgcagggggcccggtcttttgcagaccgacctgtccggtgccctgaatgaactgcag  
 gacgaggcagcgcggctatcgtggctggccacgacggcggtccttgcgcagctgtgctcgcacgttgcactgaagcgg  
 gaagggactggctgctattgggcgaagtggcggggcaggatctctgtcatctcaccttgctcctgccgagaaagtatcca  
 5 tcatggctgatgaatgcggcggtgcatacgttgatccggctacctgccattcgaccaccaagcgaacatcgcatcg  
 agcgagcagctactcggatggaagccggtctgtcgtacaggatgatctggacgaagagcatcaggggctcgcgccagc  
 cgaactgttcgccaggctcaaggcgcgcgtacccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttggcg  
 aatatcatggtggaaaatggccgctttctggattcatcgaactgtggcggctgggtgtggcggaccgctatcaggacatag  
 cgttggctacccggtgataatgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacggatcgcgcgtccc  
 10 gattcgcagcgcacatcgcttctatcgcttcttgacgagttcttgacgggactctggggttcgaaatgaccgaccaagc  
 gacgccccaacctgccatcacgagatttcgattccaccgccgcttctatgaaagggtgggcttcggaatcgtttccgggac  
 gccggctggatgatcctccagcgcggggatctcatgctggagtcttcgccaccccaactgtttattgcagcttataatggt  
 tacaataaagcaatgacatcacaaattcacaaataaagcatttttctactgcattctagtgtgtgttgcacaaactcatcaat  
 gtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttccctgtgtgaaattgtat  
 15 ccgctcacaaatccacacaacatacagagccggaagcfaaagtgtaaagcctggggtgcctaatagtgagtaactcac  
 attaatgctgttcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcg  
 gggagaggcggttgcgtattggcgctcttccgcttctcgtcactgactcgtcgcgtcggcgttcggctcggcgag  
 cggatcagctcactcaaaaggcggtataacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaa  
 ggccagcaaaaaggccaggaaccgtaaaaaggccgcgttgcgtggcgttttccataggctccgccccctgacgagcatca  
 20 caaaaatcgacgtcaagtcagaggtggcgaaacccgacaggactataaagataaccaggcggttccccctggaagctccc  
 tctgtgcgtctcgttccgacctgccggttaccggatacctgtccgcttctccttcgggaagcgtggcgcttctcaat  
 gctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccga  
 ccgctgcgccttatccggtaactatcgtctgagtcacccggtaagacacgacttatcgcactggcagcagccactggt  
 aacaggattagcagagcgaggtatgtagggcggtgctacagagttctgaagtgggtggcctaactacggctacactagaag  
 25 gacagttattggtatctgcgctcgtcgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaacca  
 ccgctggtagcgggtggtttttgttgaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatctttc  
 tacgggtctgacgctcagtggaacgaaaactcacgttaagggaattttggtcatgagattatcaaaaaggatcttcacctaga  
 tccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggtctgacagttaccaatgcttaacagtga  
 ggcacctatcagcgtatctgtctatttcgttcacatagttgctgactccccgtcgtgtagataactacgatacgggaggg  
 30 ctaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagcc  
 agccggaaggggcgcagcgcagaagtggctcctgcaactttatccgctccatccagcttattaattgttgcgggaagctag  
 agtaagtagttcggcagtaatagtttgcgcaacgttggccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatg  
 gttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtcaaaaaagcggttagctccttcgg  
 tctccgatcgttgcagaagtaagtggccgcagtggtatcactcatggtatggcagcactgcataattcttactgtcatgc  
 35 catccgtaagatgctttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctctt  
 gcccggcgtaataacgggataataccgcgccacatagcagaactttaaaagtgtcatcattgaaaacgttcttcggggc  
 gaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatctttac  
 ttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggggcgacacggaaatgt  
 tgaatactcactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagcggatcatattgaatgtatttag  
 40 aaaaataaacaatatagggttccgcgcacattccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.28: (Emerald)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 720 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

5 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

10

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....720

15 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:28:

atggtgagcaaggcgaggagctgtcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccac  
aagttcagcgtgtccggcgaggcgaggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggca  
agctgcccgtgccctggcccaccctcgtgaccacctcacctacggcgtgcagtgtctcggccgctaccccgaccacatg  
aagcagcacgacttctcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttctcaaggacgacggcaa  
20 ctacaagaccgcgccgaggtgaagttcgaggcgacaccctggtgaaccgcatcgagctgaaggcgatcgactcaa  
ggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaaggcttatatcaccgccgacaagca  
gaagaacggcatcaaggtaactcaagaccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacca  
gcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcaccagtcgccctgagcaa  
agacccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgggatcactctggcatggacgag  
25 ctgtacaagtaa

(2) INFORMATION FOR SEQ. ID. NO.29: (GFP 5' primer)

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

35

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

45

(B) LOCATION: 1.....24

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:29:  
ggatccgaattcggcaccatggtg

5

10

(2) INFORMATION FOR SEQ. ID. NO:30:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

25

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

30

(B) LOCATION: 1.....24

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:30:  
ccggaatcaaagcgcttctcagacttactt

35

2) INFORMATION FOR SEQ. ID. NO.31: (pcDNA3-1XUb-GFP)

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 6340 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45

(D) TOPOLOGY: circular



(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

5 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6340

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:31:

10 gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttggaggctcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcttgacattgattattga  
ctagtattaatagtaataattacggggcattagttcatagcccataatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
15 ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagatgacctatgggactttctacttggcagtag  
atctacgtattagtcacgctattaccatggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
20 gaaccactgcttactggcttattcgaattaatacactcactatagggagaccaagcttggtagccaccatggagatcttcg  
tgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgacaa  
gacaaggaaggcatccctcctgaccagcagaggtgatcttctgctgggaacagctggaagatggacgcaccctgtctga  
ctacaacatccagaaaagagtcaccctgcacctggactccgtctcagaggtgtgcaccacggatccgaattcggccaccat  
ggtgagcaagggcgaggagctgttcaccgggggtgtgccatcctggtcagctggacggcgacgtaaacggccacaa  
25 gttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagtcatctgcaccaccggcaag  
ctgcccgtgccctggcccaccctcgtgaccaccttcacctacggcggtgcagtgttcgcccgtaccccgaccacatgaa  
gcagcacgacttctcaagtcgccatgccgaaggctacgtccaggagcgcaccatcttctcaaggacgacggcaact  
acaagaccgcgccgaggtgaagttcgaggcgacacctggtgaaccgcacgagctgaagggcacgacttcaagg  
aggacggcaacatcctggggcacaagctggagtacaactacaacgccacaaggcttatatcaccggcgacaagcaga  
30 agaagggcatcaagggtgaacttcaagacccgccacaacatcaggagcggcagcgtgcagctcggcaccactaccagc  
agaacacccccatcggcgacggccccgtgctgctgcccgaaccactacctgagcaccagctcggcctgagcaaaag  
acccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccggcgccgggtagctctcggcatggacgagct  
gtacaagtaagtctagagggccctattctatagtgacacataatgctagagctcgtgatcagcctcactgtgcttctagt  
tgccagccatctgttgtttccccctccccgtgccttcttgacctggaagggtgacctcccactgtccttctcaataaaatg  
35 aggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggatt  
gggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctag  
ggggatccccacgcgccctgtagcggcgcatgaagcgcggcggtgtgtgtgttacgcgcagcgtgaccgctacactt  
gccagcgccctagcggcgctccttctgcttcttcccttcttctcggcacgttcgccggcttccccgtcaagctctaaatc  
ggggcatcccttaggggttccgatttagtctttacggcacctcgaaccccaaaaacttgattaggggtgatgttcacgtagt  
40 ggccatcgccctgatagacggttttgcgcccttgacgttggagtcacgttctttaatagtgactcttgttccaaactggaac  
aacactcaacctatctcggtctattctttgatttataagggttttggggatttcggcctatttggttaaaaaatgagctgattaa  
caaaaatttaacgcgaattaattctgtggaatgtgtgcagttagggtgtggaagttcccaggctccccaggcaggcagaa  
gtatgcaaagcatgcattcaattagtcagcaaccaggtgtggaaagttcccaggctccccaggcaggcagaagtatgcaa  
agcatgcattcaattagtcagcaaccatagtcggcccttaactccgccatcccgccctaactccggccaggtccggcc  
45 attctccggcccatggctgactaattttttattatgcagaggccgagggcgccctctgcctctgagctattccagaagtagtg  
aggaggctttttggaggcctaggcttttcaaaaagctccgggagcttgcataccatttccgatctgatcaagagacag  
gatgaggatcgtttcgcattgattgaacaagatggattgcacgcaggttctccggcgcttgggtggagaggctattcggcta

tgactgggcacaaacagacaatcggtgctctgatgccgccgtgtccggctgtcagcgcaggggcccgggtcttttgtc  
 aagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcggctatcgtggctggccacgacgggcgtt  
 ccttgccagctgtgctcgacgttgcactgaagcgggaaggactggctgctattgggcgaagtggcgggagcaggtatc  
 cctgtcatctaccttgcctcgcgagaaagtatccatcatggtgatgcaatgcggcggctgcatacgttgatccggcta  
 5 cctgcccattcgaccaccaagcgaacatcgcacgcagcagcactcggatggaagccggtctgtcgtatcaggat  
 gatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccagggtcaaggcgcgcacgtcccacggcga  
 ggatctcgtcgtgacctatggcgtatgcttgcgaatatcatggtggaaaatggccgctttctggattcatcgactgtg  
 gccggctgggtgtggcgaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgg  
 gctgaccgcttctcgtgctttacggatcgcgcgtcccgttcgcagcgcacgccttctatcgccttctgacgagtcttct  
 10 gagcgggactctgggggtcgaatgaccgaccaagcgcgccaacctgccatcacgagattcgattccaccgccgcct  
 tctatgaaaggttgggttcggaatcgtttccgggacgccggctggatgatcctccagcgcgggagctcatgtggagtt  
 ctccgccaccccaactgtttattgcagcttataatggttacaataaagcaatagcatcacaatttcacaaataaagcattt  
 tttcactgcattctagttgtgtgttgcctaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggc  
 gtaatcatggcatagctgtttcctgtgtgaaattgttatccgtcacaattccacacaacatacagccggaagcataaagt  
 15 taaagcctgggggtgcctaatagtgagtaactcacattaattgcgttgcgtcactgcccgtttccagtcgggaaacctgt  
 cgtgccagctgcattaatgaatcgccaacgcgcggggagagggcgttgcgtattgggcgcttctccgcttctcgtca  
 ctgactcgtcgcgtcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaatacggttatccacagaat  
 caggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggcccggttgcgt  
 gcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccggacagg  
 20 actataaagataccaggcgtttccccctgggaagctccctcgtgcgtctcctgttccgacctgcccgttaccggatacctgt  
 ccgcttcttcccttcgggaagcgtggcgcttctcattgctcacgctgtaggtatctcagttcggtgtaggtcgttcgtcca  
 agctgggctgtgtgcacgaacccccgttcagccgaccgctgcgccttatccggttaactatcgtcttgagccaacccggt  
 aagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgggtgctacagagt  
 tcttgaagtgggtggcctaactacggctacactagaaggacagtatttggtatctgcgctcgtcgaagccagttaccttcgga  
 25 aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcgggtgtttttgtttgcaagcagcagattacgcg  
 cagaaaaaaaggatctcaagaagatcctttgatctttctacgggtctgacgctcagtggaacgaaaactcacgttaaggg  
 attttggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaatcaatctaagtatatg  
 agtaaaacttggtcgtacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacatagttgcc  
 tgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccca  
 30 cgctcaccggctccagatttatcagcaataaaccagccagccggaaggccgagcgcagaagtggtcctgcaactttatc  
 cgctccatccagctctattaattgttccgggaagctagagtaagtagttcgccagttaatagtttgcgaacgttgttgcatt  
 gctacaggcatcgtggtgtcacgctcgtcgtttggtatggttcattcagctccggttcccaacgatcaaggcgagttacatg  
 atccccatgttgtgcaaaaaagcggtagctccttcggtcctccgatcgttgcagaagtaagttggccgagtggtatcact  
 catggttatggcagcactgcataattcttactgtcatgccatccgtaagatgctttctgtgactggtgagtaactcaaccaagt  
 35 cattctgagaatagtgatgcggcgaccgagttgcttggccggcgtaatacgggataataccgcgccacatagcagaa  
 ctttaaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtgt  
 aaccactcgtgcaccaactgatcttcagcatctttactttaccagcgttttgggtgagcaaaaacaggaaggcaaaat  
 gccgcaaaaaagggaataaggggcgacacggaaatgttgaatactcatacttcttcttttcaatatttgaagcatttatcag  
 ggttattgtctcatgagcggatacatattgaatgtatttagaaaaataaacaataagggttccgcgcacatttccccgaaaa  
 40 gtgccacctgacgtc

45

(2) INFORMATION FOR SEQ. ID. NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6607 base pairs

5 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

10

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

15 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6607

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:32:

20 gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtagcgcttttcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaatagtaataattacggggcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgccccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
25 ttccattgacgtcaatgggtggactatttacggtaaaactgcccaacttgccagtagcatcaagtgtatcatatgccaagtacgcc  
cctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgaccttatgggactttctacttggcagtag  
atctacgtatttagtcatcgtattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aaactccgccccattgacgcaaatgggcggttagcgtgtacgggtggaggtctatataagcagagctctctggctaactaga  
30 gaaccactgcttactggcttatcgaaattaatcagctcactataggagagaccaagcttgatcgaattcctgcagccccg  
ggggatctaccatggaatcttcgtgaagactctgactggttaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaagaggtccacctgcacctggtactccgtctcagaggtgtgcac  
cacggatctaccatggaatcttcgtgaagactctgactggttaagaccatcactctcgaagtggagccgagtacaccattg  
35 agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaagaggtccacctgcacctggtactccgtctcagaggtgtgcac  
cacggatccgaattcggcaccatgggtgagcaagggcgagggagctgttaccgggggtgtgccatcctgtgctgagctgg  
acggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctg  
aagttcatctgcaccaccggcaagctgcccgtgccctggcccacctcgtgaccaccttcacctacggcgtgcagtgttc  
40 gcccgtaccgccaccatgaagcagcacgacttctcaagtcgccatgccgaaggctacgtccaggagcgcacca  
tcttctcaaggacgacggcaactacaagaccgcgccgaggtgaagttcagggcgacacctggtgaaccgcatcga  
gctgaagggtcatcactcaaggaggacggcaacatctggggcacaagctggagtacaactacaacgccacaaggt  
ctatataccgccgacaagcagaagaacggcatcaaggtgaactcaagaccgccacaacatcagaggacggcagcgt  
gcagctcggcaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgag  
45 caccagtcggccctgagcaaaagaccccaacgagaagcgcgacacatggtcctgctggagttcgtgaccgccgccgg  
gatcactctcgcatggacgagctgtacaagtaagctagagggccctattctatagtgacctaataatgtagagctcgtg  
atcagcctcgactgtgccttctagttgccagccatctgtgttggccctccccgtgccttccttgacctggaaggtgccact

cccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgcattctattctgggggggtgggggtgggg  
 caggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatggcttctgagggc  
 gaaagaaccagctgggggctctagggggtatccccacgcgccctgtagcggcgcatlaagcgcggcggtgtgtgtggtta  
 cgcgcagcgtgaccgctacacttgccagcgccttagcggcgctcttctgcttctcccttcttctcggcacgttcgccc  
 5 gctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgccttacggcacctcgacccccaaaaaacttg  
 attaggggtgatgggtcacgtagtggggccatcgccctgatagacggttttgccttttgacgttgagtcacagttcttaatag  
 tggactcttgttccaaactggaacaacactcaaccctatctcggctctattctttgattataagggaatttggggatttcggccta  
 ttggttaaaaaatgagctgatttaacaaaaatftaacgcgaattatctgtggaatgtgtgcagttagggtgtggaaagtccc  
 caggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgaaagtccccaggctc  
 10 cccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggcccttaactccgcccattcccggcc  
 ctaactccgcccagttccgcccattctccgccccatggctgactaatttttttattatgcagaggccgagggcgccctctgc  
 ctctgagctattccagaagtagtgaggaggctttttggaggccctaggttttcaaaaagctccgggagcttgatatccat  
 tttcggatctgatcaagagacaggatgaggatcgttcgcgatgtgaacaagatggattgcacgcagggttctccggccgctt  
 ggggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgctgttccggctgtcagcg  
 15 caggggcgccccgggtcttttgcgaagaccgacctgtccgggtgccctgaatgaactgcaggacgaggcagcgcggtctatc  
 gtggctggccacgacgggcgttccttgcgcagctgtgtcgcaggtgtcactgaagcgggaagggaactgggtgctattgg  
 gcgaagtgcggggcaggatctctgtcctcaccctgtcctgccgagaaagtatccatcatggctgatgcaatgcggc  
 ggctgcatacgttgatccggctacctgcccattcgaccaccaagcgaacatcgcacgagcgacgtactcggatg  
 gaagccggcttctgtcagcaggtgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccagggtca  
 20 aggcgcgcagctcccagcggcgaggatctcgtcgtgacctatggcgatgcctgcttgcggaatatcatgggtggaatggc  
 cgcttttctggattcatcagctgtggcgggtgggtgtggcgaggcgctatcaggacatagcgttggctaccctgatattgc  
 tgaagagcttggcggcgaatgggtgaccgcttctcgtgctttacgggtatcgccgctcccattcgcagcgcacgccttc  
 tatcgcttcttgacgagttcttgcagcgggactctgggggttcgaaatgaccgaccaagcgacgcccacctgccatcacg  
 agatttcgattccaccgccccttctatgaaaggttgggttcggaatcggttccgggacggcggtggatgatctccagc  
 25 gcggggatctcatgctggagttcttcccaccccaactgtttattgcagcttataatggttacaaataagcaatagcatca  
 caaatttcacaaataaagcattttttactgcattctagtgtgtgttgcctcaactcatcaatgtatcttatcatgtctgtataccgt  
 cgaccttagctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgttatccgctcacaattccacacaacat  
 acgagccgggaagcataaagtgtaaagcctgggggtgcctaatgagttagtaactacattaattgcgttgcgctcactgcc  
 cgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatggccaacgcgcggggagaggcggttgcgtattgg  
 30 gcgctcttccgcttctcgtcactgactcgtcgcgtcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcg  
 gtaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccagggaac  
 cgtaaaaaggccgcgttgcgtggcgttttccataggctccgccccctgacgagcatcacaataacgacgtcaagtcag  
 aggtggcgaaaaccgacaggactataaagataccaggcggttccccctggaagctccctcgtgcgtctcctgttccgacc  
 ctgcccgttaccggatacctgtccgcttctccttccgggaagcgtggcgcttctcaatgtcacgctgtaggtatctcagtt  
 35 cggtgtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccaccgctgcgcttatccggttaact  
 atcgtcttgagtccaacccggttaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgagg  
 tatgtaggcggtgctacagagttcttgaagtgtggcctaactacggctacactagaaggacagtttgggtatctgcgctct  
 gctgaagccagttaccctcggaaaaagagttggtagctcttgatccggcaaaaccaccgctggtagcgggtgtgttttgg  
 tttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggctgacgctcagttg  
 40 aacgaaaactcacgttaagggttttggctcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttt  
 taaatcaatctaagtatatatgagtaacttggctgtacagttaccaatgcttaatcagtgaggcacctatctcagcagctgt  
 ctatttctgtcatccatagttgctgactccccgtcgttagataactacgatacgggagggttaccatctggccccagtgct  
 gcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggggccgagcgc  
 agaagtgtcctgcaactttatccgctccatccagcttattaattgttgcgggaagctagagtaagtagttcggcagttaat  
 45 agtttgcgcaacgttgttccattgctacaggcatcgtgggtgcacgctcgtcgttggtaggttcattcagctccggttccc  
 aacgatcaaggcgagttacatgatccccatgttgcataaaaagcggttagctccttggctcctccgatcgttgcagaagt  
 aagttggccgcagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgt

gactggtgagtactcaaccaagtcattctgagaatagtgtatcgggcgaccgagttgctcttggccggcgtaatacgggat  
aataccgcgccacatagcagaactttaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaaggatcttacc  
gctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatctttactttaccagcgtttctgggtga  
gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggggcgacacggaaatgttgaatactcatactcttctttt  
5 caatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatatagggggt  
ccgcgcacatttccccgaaaagtccacctgacgtc

10

(2) INFORMATION FOR SEQ. ID. NO:33:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 6850 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

25 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6850

30

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:33:

gacggatcgggagatctccgatccctatggcgtcgtcagtcacaaatctgctctgatgcccatagtaagccagtatctg  
ctccctgcttgtgtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgtcgttcgcgatgtacgggccagatatacgcgttgacattgattattga  
35 ctagtattataatagtaaatcaattacggggctcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgcccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggaact  
ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgacctatgggactttctacttggcagtagc  
atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatggcggttgatagcgggttgactcacgg  
40 ggatttccaagtctccacccattgacgtcaatgggagtttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaaccactgcttactggcttatcgaaattaatcagactcactatagggagaccaagcttgatcgaattcctgcagccccg  
ggggatctaccatggaaatcttcgtgaagactctgactggttaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggtgatctttgctgggaaacagctg  
45 gaagatggacgcacctgtctgactacaacatccagaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttcgtgaagactctgactggttaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaaggaaggcatccctcctgaccagcagaggtgatctttgctgggaaacagctg

gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatctcgtgaagactctgactgtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaagggaagcaccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagaggtgtgcac  
 5 cacggatccgaattcggccaccatgggtgagcaagggcgaggagctgttaccgggggtgggtcccatcctggtcagagctgg  
 acggcgacgtaaacggccacaagttcagcgtgtccggcgaggggcgaggggcgatgccacctacggcaagctgacctg  
 aagttcatctgcaccaccggcaagctgcccgtgcccggcccaccctcgtgaccaccttcacctacggcggtgcagtgttc  
 gcccgtaccggaccacatgaagcagcagcacttctcaagtccgcatgcccgaaggctacgtccaggagcgacca  
 tcttctcaaggacgacggcaactacaagaccgcgccgaggtgaagttcgaaggcgacacctggtgaaccgcatcga  
 10 gctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaaggt  
 ctatataccggcgacaagcagaagaacggcatcaaggtgaacttcaagaccggccacaacatcgaggacggcagcgt  
 gcagctcgccgaccactaccagcagaacacccccatcgccgacggccccgtgctgctgcccgaaccactacctgag  
 caccagtcggccctgagcaagaccccaacgagaagcgcgacacatggtcctgctggagttcgtgaccggcgccgg  
 gatcactctcggtatggacgagctgtacaagtaagtctagagggccctattctatagtgacacctaagttagagctcgtg  
 15 atcagcctcgactgtgcttctagtgtccagccatctgtgttggccccctccccgtgcttcttaccctgggaaggtgccact  
 cccactgtcttcttaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtgggtgggg  
 caggacagcaaggggggaggttgggaagacaatagcaggcatgctggggatgcgggtgggtctatggcttctgagggc  
 gaaagaaccagctgggggtctaggggggtatccccacgcgccctgtagcggcgccattaaagcggcggggtgtggtggtta  
 cgcgcagcgtgaccgctacacttgccagcgccctagcgccccgtccttctgcttcttcccttcttctcggcacgttcgccg  
 20 gcttccccgtcaagctctaaatcggggcaccccttaggggttccgatttagtgccttacggcacctcagccccaaaaacttg  
 attaggggtgatggtcacgtagtggggccatcgccctgatagacgggttttgcgcccttgacgttggagtccacgttcttaatag  
 tggactcttgttccaaactggaacaacactcaaccctatctcgggtctattctttgatttataagggatttggggatttcggccta  
 ttggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtc  
 caggctccccaggcaggcagaagtatgcaaagcatgcacatcaattagtcagaaccatagtcggccctaacctccgccatccggc  
 25 cccagcaggcagaagtatgcaaagcatgcacatcaattagtcagaaccatagtcggccctaacctccgccatccggc  
 cctaacctccgccagttccgccatctccgcccataggtgactaattttttattatgcagaggccgaggccgcctctgc  
 ctctgagctattccagaagtagtgaggagggtttttggaggcctaggcttttgcaaaaagctccgggagcttgatatccat  
 ttctggatctgatcaagagacaggatgaggatcggttcgcattgaaacaagatggattgcacgcaggttctccggcgctt  
 ggggtggagaggctattcggtatgactgggcacaacagacaatcggtcgtctgatccgccgtgtccggctgtcagcg  
 30 caggggcgccccgttcttttgcgaagaccgacctgtccgggtgccctgaatgaactgcaggacgaggcagcgcggtctatc  
 gtggctggccacgacggggcgttcttgcgcagctgtgctcgcaggtgtcactgaagcgggaagggaactggctgctattgg  
 gcgaagtgcggggcagggatcctgtcatctcacttgcctcctgagaaagtatccatcatggctgatgcaatgcggc  
 ggctgcatacgttgatccggctacctgcccattcgaccaccaagcgaacatcgcacgagcgagcagctactcggatg  
 gaagccggtcttgcgatcaggatgatctggacgaagagcatcagggggtcgcgccagccgaactgttcgccaggtca  
 35 agggcgcatgcccgcggcgaggatctcgtcgtgacctatggcgatgcctgcttgcgaatatcatggtggaaaatggc  
 cgcttttctggattcatcgactgtggccgggtgggtgtggcgaccgctatcaggacatagcgttggctaccctgcatattgc  
 tgaagagcttggcggcgaatgggtgaccgcttctcgtgctttacggtatcgccgctcccgttcgcagcgcatcgcttc  
 tatcgcttcttgacgagttctctgagcgggactctgggttcgaatgaccgaccaagcgacgcccacctgccatcacg  
 agatttcgattccaccggcgcttctatgaaaggttgggttcggaatcggttccgggacggcggtggatgatcctccagc  
 40 gcggggatctcatgctggagttcttcgccacccccaaactgtttattgcagcttataatggttacaaataaagcaatagcatca  
 caaatctcacaataaagcattttttactgcattctagttgtgtgttgcctcaactcatcaatgtatcttatcatgtctgtataccgt  
 cgacctctagctagagcttggcgtaatcatggctcatagctgttctgtgtgaaattgttatccgctcacaattccacacaacat  
 acgagccgggaagcataaagtgtaaagcctgggtgtcctaatagtagtgagtaactacattaattgcgttgcgctcactgcc  
 cgctttccagtcgggaaacctgtcgtgccagctgcatatgaatcgcccaacgcgggggagaggcggttgcgtattgg  
 45 gcgcttctccgcttctcgtcactgactcgtcgtcgtcgttcggctgcggcgagcgggtatcagctcactcaaaaggcg  
 gtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaac  
 cgtaaaaaggccggtgctggcggttttccataggtccgccccctgacgagcatcacaaaaatcgacgctcaagtcag

aggtggcgaaacccgacaggactataaagataccaggcggttccccctggaagctccctcgtgcgctctcctgttccgacc  
 ctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagtt  
 cgggtgtaggtcgtcgtccaagctgggctgtgtgcacgaacccccgttcagcccaccgctgcgccttatccggtaact  
 atcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgagg  
 5 tatgtaggcgggtgctacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatttggtatctgcgctct  
 gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaccaccgctggtagcgggtggtttttg  
 ttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggtctgacgctcagtgg  
 aacgaaaactcacgttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagttt  
 taaatcaatctaaagtatatatgagtaaacttggctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgt  
 10 ctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggaggggttaccatctggccccagtgt  
 gcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgc  
 agaagtggctctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagtgcgcagttaat  
 agtttgcgcaacgttggccattgctacaggcatcgtgggtgcacgctcgtcttggtagtggttcattcagctccggttccc  
 aacgatcaaggcgagttacatgatcccccatgttggtgcaaaaaagcgggttagctccttcgggtcctccgatcgttgcagaagt  
 15 aagttggccgcagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgt  
 gactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggat  
 aataccgcgccacatagcagaactttaaaagtctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttacc  
 gctgttgagatccagttcgtgtaacccactcgtgcacccaactgatcttcagcatctttactttaccagcgtttctgggtga  
 gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatgttgaatactcactcttctctttt  
 20 caatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatatgggggtt  
 ccgcgcacatttccccgaaaagtccacctgacgtc

25

## (2) INFORMATION FOR SEQ. ID. NO:34:

### (i) SEQUENCE CHARACTERISTICS:

30

(A) LENGTH: 7093 base pairs

(B) TYPE: nucleic acid

35

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

40

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

45

(B) LOCATION: 1.....7093

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:34:







ggggtggagaggctattcggctatgactgggcacacagacaatcggctgctctgatgccgccgtgtccggctgtcagcg  
 cagggggcggccgggtcttttgcagaccgacctgtccgggtgccctgaatgaactgcaggacgaggcagcgggctatc  
 gtggctggccacgacggggttccttgcgcagctgtgctcagctgtgactgaagcgggaagggactggctgctattgg  
 gcgaagtgccggggcaggatctcctgtcatctaccttgcctcgcgagaaagtatccatcatggctgatgcaatgcggc  
 5 ggctgcatacgccttgatccggctacctgcccattcgaccaccaagcgaacatcgcacgagcagcactcggatg  
 gaagccgggtctgtgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctca  
 aggcgcgcacgcccacggcgaggatctcgtcgtgacctatggcgatgcctgcttgcgaatacatggtggaaaatggc  
 cgcttttctggattcatcactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggctaccctgtatattgc  
 tgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacgggtatcggcgtcccgttcgcagcgcacgccttc  
 10 tategccttcttgacgagttcttgcagcgggactctgggggttcgaaatgaccgaccaagcgacgcccacctgccatcacg  
 agatttcgattccaccgccgcttctatgaaagggtgggcttcggaatcggtttccgggacgccggctggatgatcctccagc  
 gcggggatctcatgctggagttcttcgccaccccaactgtttattgcagcttataatggttacaataaagcaatagcatca  
 caaatttcacaaataaagcatttttctactgcattctagttgtgtttgtccaaactcatcaatgtatcttatcatgtctgtataccgt  
 cgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgtatccgctcacattccacacacat  
 15 acgagccggaagcataaagtgtaaagcctgggggtgcctaagtgtgagtaactcacattaattgcgttgcgctcactgcc  
 cgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgg  
 gcgctcttccgcttctcgtcactgactcgtcgcgtcggctgctggctgcggcgagcgggtatcagctcactcaaaaggcg  
 gtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaac  
 cgtaaaaaggccgcgttgcgtggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcag  
 20 aggtggcgaaacccgacaggactataaagataccaggcggtttccccctggaagtcctcgtcgcgtctcctgttccgacc  
 ctgccgcttaccggatacctgtccgcttctccttcgggaagcgtggcgcttctcaatgtcacgctgtaggtatctcagtt  
 cgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttatccggtaact  
 atcgtcttgagccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgagg  
 tatgtaggcgggtgtacagagttctgaagtgggtggcctaactacggctacactagaaggacagtatttggatctgcgctct  
 25 gctgaagccagttaccttcggaaaaagagttggtagctctgatccggcaaaacacaccgctggtagcgggtgtttttg  
 ttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtg  
 aacgaaaactcacgttaagggttttggctatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttt  
 taaatcaatctaaagtatatatagtaaaacttggctgtacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgt  
 ctatttcgtcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggttaccatctggccccagtgct  
 30 gcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggggccgagcgc  
 agaagtggctcctgcaactttatccgctccatccagctctattaattgttggcgggaagctagagtaagtagttcgcagttaat  
 agtttgcgaacggttggcattgctacaggcatcgtgggtgcacgctcgtcgtttggtatggcttcattcagctccggttccc  
 aacgatcaaggcgagttacatgatccccatgttgcacaaaaagcggttagctccttcggctcctccgatcgttgcagaagt  
 aagtggcgcgagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgctttctgt  
 35 gactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcggcgcaatacgggat  
 aataccgcgccacatagcagaactttaaagtgtcatcattgaaaaacgttctcggggcgaaaactctcaaggatcttacc  
 gctgttgatccagttcgtatgaacccactcgtgcaccaactgatcttcagcatctttactttaccagcgtttctgggtga  
 gcaaaaacaggaaggcaaaatgcccaaaaaagggaataaggcgacacggaaatgttgaatactcactcttctttt  
 caatattattgaagcatttatcagggttattgtctcatgacggatacatatttgaatgtatttagaaaaataaacaatagggggt  
 40 ccgcgcacatttccccgaaaagtgcacctgacgtc

## 2) INFORMATION FOR SEQ. ID. NO.35: (Caspase 3)

### 45 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 834 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

5

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....834

15

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:35:

atggagaacactgaaaactcagtgattcaaaatccattaaaaatttggaaacaaagatcatacatggaagcgaatcaatgg  
actctggaatatccctggacaacagttataaaatggattatcctgagatgggtttatgtataataattaataaagaattttcata  
aaagcactggaatgacatctcggctctggtacagatgtcgcagcaaacctcagggaacattcagaaacttgaaatatg  
20 aagtcaggaataaaaaatgatcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaag  
gagcagttttgtgtgcttctgagccatgggtgaagaagggaataatttttgaacaaatggacctgttgacctgaaaaaat  
aacaaacttttcagaggggatcgtttagaagtctaactggaaaacccaaacttttcattattcaggcctgccgtgttacaga  
actggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagtgaggccgacttcttg  
tatgcatactccacagcacctggttattattctggcgaattcaaaggatggctcctgggtcatccagtcgctttgtgccaatgct  
25 gaaacagtatgccgacaagcttgaattatgcacattcttaccgggttaaccgaaaggtggcaacagaatttgagtccttttc  
ctttgacgctacttttcagaaagaacagattccatgtattgttccatgctcacaaaagaactctattttatcactaa

2) INFORMATION FOR SEQ. ID. NO.36: (C35' primer)

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 base pairs

35 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: oligonucleotides

(ix) FEATURE:

45 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....51

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:36:

5 CGGATCCAACACTGAAAACCTCAGTGGATTCAAAATCCATTAAAAATTTGG

(2) INFORMATION FOR SEQ. ID. NO:37:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

15

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

20

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

25

(B) LOCATION: 148

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:37:

30

CGGATCCGTGATAAAAATAGAGTTCTTTTGTGAGCATGGAAACAATAC

35

2) INFORMATION FOR SEQ. ID. NO.38: (pcDNA3-1XUb-C3)

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 6436 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5 (B) LOCATION: 1.....6436

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:38:

gacggatcgggagatctcccgatcccctatggcgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctgctgagtagtgcgcgagcaaaatctaagctacaacaaggcaaggctgaccgacaatt  
10 gcatgaagaatctgcttagggtagggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaatagtaataaattacggggcattagttcatagcccataatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
cctattgacgtcaatgacgtaaatggcccgcctggcattatgccagtagacgttatgggactttcctacttggcagtag  
15 atctacgtattagtcacgcgtattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccaccccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aaactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaaccactgcttactggcttaccgaaattaatacgaactcactatagggagacccaagcttgggtaccaccatggagatcttcg  
tgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgatccaa  
20 gacaagggaaggcatccctcctgaccagcagaggtgatcttctgctgggaaacagctggaagatggacgcaccctgtctga  
ctacaacatccagaaagagtgccacctgcacctggtactccgtctcagaggtgtgcaccacggatccaacactgaaaactc  
agtggattcaaaatccattaaaaatttgaaccaaagatcatatgggaagcgaatcaatggactctggaatatccctggaca  
acagttataaaatggattatcctgagatgggtttatgtataataataataagaattttcataaaagcactggaatgacatctc  
ggtctggtagacatgtcgtatgcagcaaacctcagggaacattcagaaactgaaatatgaagtcaggaataaaatgatct  
25 tacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggagcagttttgttgtgtcttctg  
agccatgggtgaagaaggaataattttggacaaatggacctgttgacctgaaaaaaataacaaacttttcagaggggatc  
gtttagaagtctaactggaaaacccaaactttcattattcaggcctgccgtggtacagaactggactgtggcattgagaca  
gacagtgggtgtgatgatgacatggcgtgtcataaaataaccagtggaggccgacttctgtatgcatactccacagcacctg  
gttattattcttggcgaaattcaaggatggctcctggttcacagtcgcttgtgccatgtgaaacagtagccgacaagct  
30 tgaatttatgcacattcttaccgggttaaccgaaaggtggcaacagaatttgagtccttttctttgacgtacttttcatgcaa  
agaaacagattccatgtattgtttccatgctcacaaaagaactctattttatcacggatcctagaggccctattctatagtgc  
acctaaatgctagagctcgtgatcagcctcgaactgtgccttctagttgccagccatctgttgtttgccctccccgtgccttc  
cttgacctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattct  
attctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtctggggatgcggt  
35 gggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatcccccacgcgcctgtagcggcgcattaa  
gcgcggcgggtgtgtgtgttacgcgcagcgtgaccgctacacttgcagcgccttagcggccgctcttctcttcttcc  
cttcttcttccgccagttcggcggtttccccgtcaagctctaaatcggggcatcccttaggggttccgatttagtgctttacgg  
cacctcgaacccaaaaaacttgattaggggtgatggttcacgtagtgggccatcgcctgatagacggtttccgccccttgac  
gttggagtcacgttcttaataagtggactctgttccaaactggaacaacactcaacctatctcggctattcttttgattataa  
40 gggattttggggatttcggcctattgttaaaaaatgagctgatttaacaaaatlaacgcgaattattctgtggaatgtgtgt  
cagttaggggtgtggaagtccccagggtccccaggcaggcagaagtatgcaaaagcatgcatctcaattagtcagcaacca  
gggtgtggaagtccccagggtccccagcaggcagaagtatgcaaaagcatgcatctcaattagtcagcaaccaatagtcagg  
cccctaactccgccatcccgcccctaactccgccagttccgccattctccgcccatggctgactaatttttttttattatgc  
agaggccgagccgctctgctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaag  
45 ctcccgggagcttgcataatcatttccggtatgcataagagacaggatgaggatcgtttcgcatgattgaacaagatggattg  
cacgcagggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgcc  
gccgtgtccggctgtcagcgcagggggcggcggttcttttgcgaagaccgacctgtccgggtgcctgaatgaactgcag

gacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgactgaagcgg  
 gaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatcca  
 tcatggctgatgcaatgcggcggctgcatacgttgatccggctacctgccattcgaccaccaagcgaaacatcgcatcg  
 agcgagcagctactcggatggaagccggcttctgctgatcaggatgatctggacgaagagcatcaggggctcgcgccagc  
 5 cgaactgttcgccaggctcaaggcgcgcgatccccgacggcgagatctcgtcgtgacccatggcgatgcctgcttgccg  
 aatatcatggtggaatggccgctttctggattcatcactgtggccggctgggtgtggcggaccgctatcaggacatag  
 cgttggctaccgctgataatgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacggtatcgccgctccc  
 gattcgcagcgcacgccttctatgccttcttgacgagttcttctgagcgggactctggggttcgaatgaccgaccaagc  
 gacgccaacctgccatcacgagatttcgattccaccggcgccttctatgaaagggtgggcttcggaatcggtttccgggac  
 10 gccggctggatgatctccagcgcggggatctcatgctggagtcttcgccacccaacttggttattgcagcttataatggt  
 tacaataaagcaatagcatcacaaatttcacaaataaagcatttttctactgcattctagtgtggtttgccaaactcatcaat  
 gtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatacatggtcatagctgttccctgtgtgaaattgtat  
 ccgctcacaaatccacacaacatacagcgggaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactcac  
 attaatgcttgcgctcactgcccgcttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcg  
 15 gggagaggcgggttgcgtattgggcgctcttccgcttccctgcctcactgactcgtcgcctcggctggtcggctgcggcgag  
 cggtatcagctcactcaaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaa  
 ggccagcaaaaggccaggaaccgtaaaaaggccgcttgcgtggcgttttccataggtccgccccctgacgagcatca  
 caaaaatcgacgctcaagtcagagggtggcgaaacccgacaggactataaagataaccaggcggttccccctggaagctccc  
 tctgtcgcctcctgttccgacctgccgcttaccggatacctgtccgccttctccctcgggaagcgtggcgcttctcaat  
 20 gctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagccga  
 ccgctgcgccttatccgtaactatcgtcttgagtccaacccggtaagacacgactatcgccactggcagcagccactggt  
 aacaggattagcagagcggaggtatgtaggcgggtgtacagagttcttgaagtgggtgcctaactacggctacactagaag  
 gacagtatttggatctgcgctcgtcgaagccagttaccttcgaaaaagagttggtagctcttgatccggcaaacaaacca  
 ccgctggtagcgggtggtttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttc  
 25 tacggggctgacgctcagtggaacgaaaactcacgttaagggtatttggctatgagattatcaaaaaggatcttcacctaga  
 tccttttaataaaaaatgaagtttaataatctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaacagtga  
 ggcacctatctcagcgaatctgtctatttcgttcacatagttgcctgactccccgtcgtgtagataactacgatacgggaggg  
 cttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaataaaccagcc  
 agccggaaggccgagcgcagaagtggtcctgcaactttatccgctccatccagcttattaattgttccgggaagctag  
 30 agtaagtagttccagttaatagtttgcgcaacgttggccatttgctacaggcagcgtggtgtcacgctcgtcgtttggtatg  
 gcttcattcagctccggttcccaacgatcaaggcgaagtacatgatccccatgttgtgcaaaaaagcggttagctccttcgg  
 tcctccgatcgttgcagaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataattcttactgtcatgc  
 catccgtaagatgcttttctgtactggtgagtactcaaccaagtcattctgagaatagtgtatcgggcgaccgagttgctctt  
 gcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggc  
 35 gaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatctttac  
 ttaccagcgtttctgggtgagcaaaaacaggaaggcgaataaggcgacacgggaatgt  
 tgaatactcactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgtatttag  
 aaaaataaacaataagggttccgcgcacatttccccgaaaagtgcacactgacgtc

40

(2) INFORMATION FOR SEQ. ID. NO:39:

(i) SEQUENCE CHARACTERISTICS:

45

(A) LENGTH: 6703 base pairs

(B) TYPE: nucleic acid



aggggtccgatttagtgctttacggcacctcgacccccaaaaaacttgattaggggtgatgggtcacgtagtgggccatcgccct  
 gatagacgggttttcgccctttgacgttggagtcacgttcttaatagtggactcttgttccaaactggaacaacactcaaccc  
 tatctcgggtctattctttgattataagggattttggggatttcggcctatttggttaaaaaatgagctgatttaacaaaaatthaac  
 gcgaattaattctgtggaatgtgtgtcagttaggggtgtgaaagtccccaggtccccaggcaggcagaagtatgcaaaagc  
 5 atgcatctcaatttagtcagcaaccaggtgtggaagtccccaggtccccagcaggcagaagtatgcaaaagcatgcatctc  
 aattagtcagcaaccatagtcctcgcccttaactccgcccatactccgccccttaactccgcccagttccgcccattctccgccc  
 atggctgactaatttttttattatgcagaggccgagggcgctctgctctgagctattccagaagttagtgaggaggctttttt  
 ggaggccataggttttgcaaaaagctccccgggagcttgtatataccattttcggatctgatcaagagacaggatgaggatcggt  
 tcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcaca  
 10 acagacaatcggctgctctgatgccgctgtgtccggctgtcagcgcaggggcgccgggtctttttgcaagaccgacct  
 gtccgggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccctgcgcagct  
 gtgctcgacgttgcactgaagcgggaaggaggactggctgctattgggcgaagtccggggcaggatctcctgtcatctcac  
 cttgctcctgccgagaagtatccatcatggctgatgaatgcggcggtgcatacgcttgatccggctacctgcccattcg  
 accaccaagcgaaacatcgatcgagcgagcagctactggatggaagccggcttctgcatcaggatgatctggacga  
 15 agagcatcaggggctcgccagccgaactgttcgcccaggtcgaagcgcgcatgccgacggcgaggatctcgtcgt  
 gacccatggcgatgcctgcttgcggaatatcatggtggaagtggcgcttttctggattcatcactgtggccggctgggt  
 gtggcgagaccgctatcaggacatagcgttggctaccgctgataattgctgaagagcttggcgagcgaatgggtgacggcttc  
 ctgctgctttacggatcgccgctcccgattcgagcgcacgcttctatcgcttcttgacgagttcttctgagcgggactct  
 ggggttcgaaatgaccgaccaagcgacgcccacctgccatcacgagattcgattccaccgccgcttctatgaaagggt  
 20 gggcttcggaatcggtttccgggacggcggtggatgatcctccagcgcggggatctcatgctggagtcttcgccaccc  
 caactgtttattgcagcttataatggttacaaataaagcaatagcatcacaatttcacaaataaagcattttttcactgcattct  
 agttgtggtttgccaactcatcaatgtatcttcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtc  
 atagctgtttctgtgtgaaattgttatccgctcacaattccacacacatacagccggaagcataaagtgtaaagcctggg  
 gtgcctaagtgtgagtaactcacattaattgcgttgcgctcactgcccgccttcagtcgggaaacctgtcgtgccagctg  
 25 catfaatgaatcgccaacgcgcggggagaggcggtttgcgtattggcgctcttccgcttctcgtcactgactcgtcgc  
 gctcggtcgttcggctcgggcgagcggtatcagctcactcaaaggcggtatacggttatccacagaatcaggggataac  
 gcaggaaagaacatgtgagcaaaaaggccagcaaaaaggccagggaaccgtaaaaaggccgcgttgcgtggcggttttccata  
 ggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccgacaggactataaagata  
 ccaggcggtttccccctggaagctccctgtgcgctcctgttccgacctgccgcttaccggataacctgtccgcttttctccc  
 30 ttcgggaagcgtggcgcttttcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgt  
 gcacgaacccccgttcagcccagcgctgcgccttatccggttaactatcgtcttgagtccaacccggtaagacacgactt  
 atcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtgg  
 ggcttaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaataagagtgg  
 agctcttgatccggcaacaaaccaccgctggtagcggtgtttttgtttgcaagcagcagattacgcgcagaaaaaag  
 35 gatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatga  
 gattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttgg  
 ctgacagttaccaatgcttaacagtgaggcacctatctcagcgtctgtctatttcgttcatccatagttgcctgactccccgtc  
 gtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc  
 tccagatttatcagcaataaaccagccagccgggaaggccgagcgcagaagtggctcctgcaactttatccgctccatcca  
 40 gtctattaattgttgccgggaagctagagtaagtagttcgccagttatagtttgcgaacgttgttgccattgctacaggcatc  
 gtggtgtcacgctcgtcgtttggtatggcttcattcagctccgggtcccaacgatcaaggcgagttacatgatccccatgttg  
 tgcaaaaaagcggttagctccttcggtcctccgatcgttgcagaagtaagttggccgcagtgttatcactcatggttatggca  
 gcaactgcataattcttactgtcatgccatccgtaagatgctttctgtgactggtgagtagtactcaaccaagtcattctgagaata  
 gtgtatcgggcgaccgagttgctcttgcggcgctcaatcgggataataccgcgccacatagcagaactttaaaagtgt  
 45 catcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
 caccacactgatcttcagcatcttttacttaccagcggttctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaaa  
 ggggaataaggcgacacggaaatgtgaatactcatacttcttcttcaatattattgaagcatttatcagggttattgtctcat

gagcggatacatatttgaatgtatttagaaaaataaacaatataggggttccgcgcacattccccgaaaagtgccacctgac  
gtc

5 (2) INFORMATION FOR SEQ. ID. NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6946 base pairs

10

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

20

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6946

25 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:40:

gacggatcgggagatctcccgatcccctatggcgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgctgtgtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagtattataatagtaaatcaattacggggctcattagttcatagcccatatatggagtccgcgttacataacttacggtaaaggc  
30 ccgctggtgacgcccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggaact  
ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
atctacgtattagtcacgctattaccatgggtgatgcggtttggcagtagcatcaatggcggtgtagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
35 aactccgeccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggttaactaga  
gaaccactgcttacttggttatcgaaattaatacagactcactataggagaccaagcttgatcgaattcctgcagcccg  
ggggatctaccatggaaatcttctgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
40 cacggatctaccatggaaatcttctgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttctgaagactctgactggtgaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
45 gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
cacggatccaacactgaaaactcagtgattcaaaatccattaaaaatttgaaccaaagatcatacatggaagcgaatcaa  
tggaactctggaatatccctggacaacagttataaaatggattatcctgagatgggtttatgtataataataataagaatttc



ataaaaactggaatgacatctcggctcgtacagatgctgatgcagcaaacctcagggaacattcagaaacttgaaata  
 tgaagtcagggaataaaaatgatcttacacgtgaagaaatttggaattgatgcgtgatgttctaagaagatcacagcaaaa  
 ggagcagttttgtgtgcttctgagccatggtgaagaagggaataattttggaacaaatggacctgttgacctgaaaaaa  
 taacaaacttttcagaggggagcgtgttagaagttaactggaaaacccaaacttttcattatcaggcctgccgtgttacag  
 5 aactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagtgagggccgacttctt  
 gtatgcatactccacagcacctggttatttcttggcgaattcaaaaggatggctcctggttcacagtcgctttgtgccatg  
 ctgaaacagtatgccgacaagcttgaatttatgcacattcttaccggggttaaccgaaagggtggcaacagaatttgagtcctt  
 tcctttgacgctacttttcacaaagaacacagattccatgtattgtttccatgctcacaagaactctattttatcacggatcct  
 agagggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgt  
 10 tgtttgccccctccccgtgccttcttgacctggaagggtgccactcccactgtcctttcctaataaaaatgaggaaattgcac  
 gcattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaata  
 gcaggcatgctggggatgagggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccca  
 cgcgccctgtagcggcgcaataagcgcggcggtgtggtggttacgcgcagcgtgaccgtacacttgccagcgcccta  
 gcgcccgtcctttcgtttcttcccttcttccgccacgtcggcggttccccgtcaagctctaaatcggggcatccctt  
 15 aggggtccgatttagtgccttacggcacctcgaccccaaaaacttgattagggtgatggttcacgtagtggggccatcgccct  
 gatagacggttttcgcctttgacgttggagtcacgttcttaatagtggactcttgtccaaactggaacaacactcaaccc  
 tatctcggcttattctttgattataagggtatttggggatttcggcctattggttaaaaatgagctgatttaacaaaaattaac  
 gcgaattaattcttggaatgtgtgcagttagggtgtggaaagtcgccaggtccccaggcaggcagaagtatgcaaaagc  
 atgcatctcaattatgcagcaaccagggtgtggaagtcgccaggtccccagcaggcagaagtatgcaaaagcatgcatctc  
 20 aattagtcagcaaccatagtcgccgccctaacctccgcccatcccgccctaacctccgccagttccgccattctccgcccc  
 atggctgactaattttttttatgcagaggccgaggccgctctgctctgagctattccagaagtatgaggaggcttttt  
 ggaggcctaggttttgcaaaaagctcccgggagcttgcataccatttcggatctgatcaagagacaggatgaggatcgtt  
 tcgcatgattgaacaagatggattgcacgcagggttctccggcgcttgggtggagaggctattcggctatgactgggcaca  
 acagacaatcggctgctctgatgccgccgtgtccggctgtcagcgcaggggcgcccggttcttttgcagaccgacct  
 25 gtccggtgccccgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggcgcttcttgcgcagct  
 gtgctgcaggtgtcactgaagcgggaagggaactggctgctattggcggaagtccggggcaggatcctctgcatctcac  
 ctgctcctgccgagaagatccatcatggctgatgcaatgcggcggtgcatatcgcttgatccggctacctgcccattcg  
 accaccaagcgaacatcgcatcgagcagcagcgtactcggatggaagccggcttctgcatcaggatgatctggacga  
 agagcatcagggggtcgcgccagccgaactgttcgccaggctcaaggcgcgcagatcccgacggcgaggatctcgtcgt  
 30 gacctatggcgatgcctgcttgcgaatatcatggtggaaatggcgcttttctggattcatcactgtggcgggtgggt  
 gtggcgaggaccgtatcaggacatagcgttgctaccgtgatattgctgaagagcttggcggcgaatgggctgaccgctt  
 ctcgtgctttacggatcgcgcctccgattcgagcgcacgccttctatgccttctgacgagttcttctgagcgggactct  
 ggggttcgaaatgaccgaccaagcgacgcccacctgccatcacgagatttcgattccaccgcccgttctatgaaaggtt  
 gggcttcggaatcgtttccgggacggcggtggtgatccctcagcgcggggatctcatgctggagttcttcccaccc  
 35 caactgtttattgcagctataatggttacaaataaagcaatagcatcacaatttcacaaataaagcattttttcactgcattct  
 agttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgaccttagctagagcttggcgtaatcatggtc  
 atagctgttctctgtgtgaaattgttatccgctcacaattccacacacatacagccggaagcataaagtgtaaagcctggg  
 gtgcctaataagtgagtaactcacattaattgcgttgcgtcactgcccgtttccagtcgggaaacctgtcgtgccagctg  
 cattaatgaatcggccaacgcgcggggagaggcggttgcgtattggcgcttctccgcttctcgtcactgactcgtgc  
 40 gctcggctgttcggctgcggcgagcggatcagctcactcaaaggcggtataacggttatccacagaatcaggggataac  
 gcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccggttgcgtggcgtttttccata  
 ggctccgccccctgacgagcatcacaataacgacgctcaagtcagaggtggcgaaaccgacaggactataaagata  
 ccaggcggtttccccctggaagctccctcgtgcgtctcctgttccgacctgccgcttaccggatactgtccgcctttctccc  
 ttccgggaagcgtggcgcttttcaatgctcacgctgtaggtatctcagtcggtgtaggtcgttcgctccaagctgggctgtgt  
 45 gcacgaacccccgttcagccgaccgtcgccttatccggttaactatcgtcttgagccaacccggttaagacacgactt  
 atcgccactggcagcagccactggttaacaggattagcagagcgaggtatgtaggcggtgtacagagttcttgaagtgtt  
 ggccctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaagaggttgg

agctcttgatccggcaacaaccaccgctgtagcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaag  
 gatctcaagaagatccttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgftaagggttttggtcatga  
 gattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagttttaaatcaatctaaagtatatagtaaacttgg  
 ctgacaggttaccatgcttaacagtgaggcacctatctcagcgatctgtctatttcgttcacatagttgcctgactccccgtc  
 5 gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc  
 tccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggctcctgcaactttatccgcctccatcca  
 gtctattaattgttgccgggaagctagagtaagtagttccaggttaataagtttgcgcaacgttggccattgtacaggcatc  
 gtggtgtcacgctcgtcgtttggtatggcttcattcagctccgggttccaacgatcaaggcgagttacatgatccccatgttg  
 tgcaaaaaagcgggttagctcctcggctcctccgatcgttgtcagaagtaagttggccgagtggtatcactcatggttatggca  
 10 gcaactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaata  
 gtgtatcgggcgaccgagttgctcttgccccggcgtaatacgggataataccgcgccacatagcagaactttaaagtgtc  
 catcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
 cacccaactgatcttcagatcttttactttaccagcggttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaa  
 gggaataaggcgacacggaaatgtgaatactcatacttcttttcaatattattgaagcatttatcagggttattgtctcat  
 15 gagcggatacatatttgatgtatttagaaaaataaacaatatgggggtccgcgcacattccccgaaaagtgccacctgac  
 gtc

## (2) INFORMATION FOR SEQ. ID. NO:41:

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7189 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

### (ii) MOLECULE TYPE: cDNA

### (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7189

### (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:41:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcagtaagccagtatctg  
 ctccctgcttggtgtggaggtcgtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
 gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
 ctagttaataatagtaataacacgaggggtcattagttcatagcccatataggagttccgcgttacataacttacggttaaattggc  
 ccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggaact  
 40 ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
 cctattgacgtcaatgacggtaaaatggcccgctggcattatgccagtagcatgacattatgggactttcctacttggcagtag  
 atctacgtattagtcacgctattaccatggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg

ggatttccaagtctccacccattgacgtcaatgggagttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
 gaaccactgttactggcttactgaaattaatacgaactactataggagacccaagcttgatacgaattcctgcagcccg  
 ggggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 5 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagagggtgtgcac  
 cacggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagagggtgtgcac  
 10 cacggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagagggtgtgcac  
 cacggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagagggtgatctttgctgggaaacagctg  
 15 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggactccgtctcagagggtgtgcac  
 cacggatccaacactgaaactcagtggtattcaaaatccattaaaaatttgaaccaaaagatcatacattggaagcgaatcaa  
 tggactctggaataccctggacaacagttataaaatggattatcctgagatgggtttatgtataataataataagaatttct  
 ataaaagcactggaatgacatctcgtctggtacagatgtcgtatgcagcaaacctcagggaaacattcagaaacttgaata  
 tgaagtcaggaataaaaaatgatcttacacgtgaagaaatttggaaattgatgcgtgatgtttctaaagaagtcacagcaaaa  
 20 ggagcagttttgtgtgtcttctgagccatggtgaagaaggaaataatttttgaacaaatggacctgttgacctgaaaaaaa  
 taacaaacttttccagaggggagctgttagaagttaactgaaaacccaaacttttattattcagccctgccgtggtacag  
 aactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagtgaggccgacttctt  
 gtatgcatactccacagcaccctggttattattcttggcgaattcaaaaggatggctcctggttcacagtcgctttgtgccatg  
 ctgaaacagtatgccgacaagcttgaatttatgcacattcttaccgggttaaccgaaagtgaggcaacagaatttgagtccttt  
 25 tctttgacgctacttttcatgcaaagaacagattccatgtattgtttccatgctcacaaaagaactctattttatcacggatcct  
 agagggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctcactgtgcccttctagtgtccagccatctgt  
 tgtttgccccccccgtgcccttctgacctggaagggtgccactcccactgtcccttcttaataaaatgaggaaattgcatc  
 gcattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaata  
 gcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatcccca  
 30 cgcgccctgtagcggcgcattaagcgcggcgggtgtggtgttacgcgcagcgtgaccgtacacttgccagcgcctta  
 gcgcccgtcctttcgtttcttcccttcttctgccacgttcgccggcttccccgtcaagctctaaatcggggcatcccttt  
 aggggtccgatttagtgctttacggcacctcgaacccaaaaaacttgattaggggtgatggttcacgtatgggccatcgcctt  
 gatagacgggttttcgccctttgacgttggagtcacgttcttaatagtgactctgttccaaactggaacaacactcaaccc  
 tatctcggctctattctttgattataagggattttggggatttcggcctattggttaaaaaatgagctgatttaaaaaaatttaac  
 35 gcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagc  
 atgcatctcaattagtcagcaaccagggtgtgaaagtccccaggctccccaggcaggcagaagtatgcaaagcagtcacatc  
 aattagtcagcaaccatagtcgccgccctaactccgccatccccgccctaactccgccagttccgccattctccgcccc  
 atggctgactaattttttattatgcagaggccgaggccgcctctgacctatccagaagtagtgaggagggttttt  
 ggaggccataggcttttcaaaaagctccgggagcttgcataccatttccgatctgatcaagagacaggatgaggatcgtt  
 40 tcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcaca  
 acagacaatcggctgctctgatgccgccgtgtccggctgtcagcgcaggggcgcccggttcttttgcagaccgacct  
 gtccgggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccctgcgcagct  
 gtgctcgcaggtgtcactgaagcgggaaggactggctgctattggcggaagtccggggcaggatctcctgtcattcacc  
 cttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgttgatccggctacctgccattcg  
 45 accaccaagcgaacatcgcacgcagcagcgtactcggatggaagccggtcttgcgatcaggatgatctggacga  
 agagcatcaggggctcgcgccagccgaactgtcggcaggctcaaggcgcgcagatcccgacggcgaggatctcgtcgt  
 gacctatggcgatgcctgcttgcgaatatcatggtggaatggccgcttttctggattcatcactgtggccggctgggt

gtggcggaccgctatcaggacatagcgttgggtacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttc  
ctcgtgctttacgggtatcgccgctcccgttcgcagcgcacgccttctatcgcttcttgacgagttctctgagcgggactct  
gggggtcgaatgaccgaccaagcgacgcccacctgccatcacgagatttcgattccaccgccgcttctatgaaaggtt  
5 ggggttcggaatcgttttcgggacgccgggtggatgacccaccgagcggggatctcatgctggagttcttcgccaccc  
caacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttactgcattct  
agttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtc  
atagctgtttctgtgtgaaattgttatccgctcacaaattccacacaacatacagagccggaagcataaagtgtaaagcctggg  
gtgcctaataagtgagctaaactcacatttaattgcgttgcgctcactgccgcttccagtcgggaaacctgtcgtgccagctg  
cattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggcgctcttccgcttctcgtcactgactcgtcgc  
10 gctcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcggtatacgggtatccacagaatcaggggataac  
gcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcggttttccata  
ggctccgccccctgacgagcatcacaaaaatcgacgtcaagtcagagtggtgcgaaaccgacaggactataaagata  
ccaggcggttccccctggaagctccctcgtgcgtctctgttccgacctgccgcttaccggatacctgtccgcttttctccc  
ttcgggaagcgtggcgcttttctaatgtctcacgctgtaggtatctcagttcgggtgtaggtcgttcgtccaagctgggctgtgt  
15 gcacgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttgagccaaccggtaagacacgactt  
atcgccactggcagcagccactggttaacaggattagcagagcgaggtatgtaggcgggtgtacagagttctgaagtgg  
ggcctaactacggctacactagaaggacagtatttggatctgcgctcgtctgaagccagttaccttcgaaaaagagtgg  
agctcttgatccggcaaacaaaccaccgctggtagcgggtgtttttgttgaagcagcagattacgcgcagaaaaaag  
gatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttggcatga  
20 gattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagttttaaatcaatctaaagtatatatgagtaacttgg  
ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgtatctgtctatttcgttcacatagttgcctgactccccgtc  
gtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc  
tccagattatcagcaataaaccagccagccggaaggccgagcgcagaagtggctcctgcaactttatccgctccatcca  
gtctattaattgttccgggaagctagagtaagtgttcgccagttaatagtttgcgcaacgttggtgccattgctacaggcatc  
25 gtggtgtcacgctcgtcgtttgtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttg  
tgcaaaaaagcggttagctccttcggctcctccgatcgttgatcagaagtaagttggccgcagtggtatcactcatggttatggca  
gcactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaata  
gtgtatgcggcgaccgagttgctcttgcggcggtcaatacgggataataccgcgccacatagcagaactttaaaagtgt  
catcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg  
30 caccacactgatcttcagcatctttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa  
gggaataaggggcgacacggaaatgtgaatactcactcttcttttcaatattattgaagcattatcagggttattgtctcat  
gagcggatacatatttgaatgtatttagaaaaataacaaataggggtccgcgcacatttccccgaaaagtgccacctgac  
gtc

35

2) INFORMATION FOR SEQ. ID. NO.42: C35'Met primer

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 54 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

45

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

5 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....53

10 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:42:

CGGATCCATGAACACTGAAAACCTCAGTGGATTCAAAATCCATTAAAAATTTGG

15

2) INFORMATION FOR SEQ. ID. NO.43: C33' primer

(i) SEQUENCE CHARACTERISTICS:

20

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

30

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

35 (B) LOCATION: 148

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:43:

40

CGGATCCGTGATAAAAATAGAGTTCTTTTGTGAGCATGGAAACAATAC

2) INFORMATION FOR SEQ. ID. NO.44: pcDNA3-Ub-Met-C3

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7248 base pairs



cattgcagcactggggccagatggtaagccctcccgtatcgtagtattctacacgacggggagtcaggcaactatggatga  
 acgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaattctagagggccctattctatagtgtcaccta  
 aatgctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgttgtttgcccctccccgtgccttccctga  
 ccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgcattctattct  
 5 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggc  
 tctatggcttctgagggcggaaagaaccagctgggggtctaggggtatccccacgcgccctgtagcggcgcatgaagcgc  
 ggggggtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccctagcggcgctccttctgctttctcccttct  
 ttctgccacgttcggcggtttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacct  
 cgacccccaaaaaactgattagggtgatggttcacgtagtggccatcgcctgtatagacggttttcgcctttgacgttg  
 10 agtccacgttcttaatagtggactctgttccaaactggaacaacactcaaccctatctcggctctattctttgattataaggga  
 ttttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatfaacgcgaattaattctgtggaatgtgtgcagt  
 aggggtgtggaagtccccagggtccccaggcaggcagaagtatgcaaagcatgcacatcaattagtcagcaaccagggtgt  
 ggaaagtccccagggtccccagcaggcagaagtatgcaaagcatgcacatcaattagtcagcaaccatagtcggccct  
 aactccgcccattcccggcctaactccggcagttccggccttctccggccatggctgactaattttttattatgcagag  
 15 gccgagggcgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctagggttttgcaaaaagctcc  
 cgggagcttctatatccatttccgatctgatcaagagacaggatgaggatcgtttcgcagattgaacaagatggattgcac  
 gcaggttctccggcgcttgggtggagaggctattcggctatgactgggcacacagacaatcggctgctctgatccgc  
 cgtgttccggctgtcagcgcagggggcgccgggttcttttgcgaagaccgacctgtccgggtgccctgaatgaactgcagga  
 cgaggcagcgcggctatctgtggtggccacgacgggcgttcttgcgcagctgtgctcgcaggtgtcactgaagcggga  
 20 agggactggctgtattggcggaagtgcggggcaggatctctgtcatctacattgctcctgccgagaaagtatccatca  
 tggctgatgcaatgcggcggtcgcatacgttgatccggctacctgccattcgaccaccaagcgaacatcgcacgag  
 cgagcacgtactcggatggaagccggcttctgtcgcacaggatgatctggacgaagagcatcaggggctcgcggcagccg  
 aactgttcggcaggtcaaggcgcgcacgcccgcggcgaggatctcgtcgtgacctatggcgatgcctgcttgcggaat  
 atcatggtggaaaatggcgcttttctggattcactgactgtggcggtggtgtggcgaccgctatcaggacatagcgt  
 25 tggctacccgtgatattgctgaagagcttggcggcgaatggggtgaccgcttctcgtgctttacgggtatcggcgctcccgat  
 tcgcagcgcacgccttctatcgccttctgacgagttctctgagcgggactctgggggtcgaatgaccgaccaagcgac  
 gcccacctgccatcacgagatttcgattccaccgcccgttctatgaaaggttgggcttcggaatcgtttccgggacgcc  
 ggctggatgatctccagcgcggggatctcatgctggagtcttccgccaccccaactgtttattgcagcttataatggttac  
 aaataaagcaatagcatcacaatttcacaaataaagcatttttactgcattctagtgtgttgcctaaactcatcaatgta  
 30 tcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgtaaatgttatccg  
 ctcaaatccacacaacatacagccggaagcataaagtgtaaagcctgggggtgcctaatgagtgcagtaactacatta  
 attgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgcccaacgcgcgggg  
 agaggcggttgcgtattggcgctcttccgcttctcgtcactgactgctcgctcggctggtcggctgcggcgagcgg  
 taccagctcactcaagggcgttaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc  
 35 cagcaaaaggccaggaaccgtaaaaaggccggttgcgtggcgttttccataggctccggcccccgtgacgagcatcaca  
 aaatcgacgctcaagtcagaggtggcgaacccgacaggactataaagataccaggcgtttccccctggaagctccctcg  
 tgcgtctctgttccgacctgccgcttaccggatacctgtccgcttctcccttcgggaagcgtggcgctttctcaatgct  
 cacgctgtaggtatctcagttcgggtgtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccagccg  
 ctgcgcttatccggtaactatcgtctgagtcacccggtaagacacgacttatcggcactggcagcagccactggtaac  
 40 aggattagcagagcgaggtatgtaggcgggtgtacagagtcttgaagtgggtggcctaactacggctacactagaaggac  
 agtatttggatctgcgctctgctgaagccagttaccttcgaaaaagagttggtagctcttgatccggcaacaaccaccg  
 ctggtagcgggtgttttttgttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctac  
 ggggtctgacgctcagtggaaacgaaactcacgttaagggttttggcatgagattatcaaaaaggatcttcacctagatcc  
 ttttaaatataaaatgaagtttaaatcaatctaaagtatatagtaaaacttggctgacagttaccaatgcttaacagtgagg  
 45 cacctatctcagcgtatctgtctatttcgttaccatagttgcctgactccccgtcgtgtagataactacgatacgggagggtt  
 accatctggccccagtgtgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccag  
 ccggaagggccgagcgcagaagtgggtcctgcaactttatccgctccatccagtctattaattgttccgggaagctagagt

aagtagttcgccagttaatagtttgcgcaacgttggtgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggct  
 tcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttggtgcaaaaaagcggtagctccttcggtcc  
 tccgatcgttgtagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccat  
 ccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcc  
 5 cggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggcgaa  
 aactctcaaggatcttacgcgtgttgagatccagttcgtatgaaccactcgtgcacccaactgatcttcagcatctttactttc  
 accagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatgtga  
 atactcactcttcttttcaatattattgaagcattatcagggttattgtctcatgagcggatacatattgaatgtatttagaaa  
 aataaacaatataggggtccgcgcacatttccccgaaaagtgccacctgacgtc  
 10

## 2) INFORMATION FOR SEQ. ID. NO.45: DEVD-1 primer

### 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

### 25 (ii) MOLECULE TYPE: oligonucleotide

#### (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

30

(B) LOCATION: 1....48

#### (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO: 45:

35 GATCCGTCGGCGCTGTCTGGCAGCGTCGGCGACGAGGTCTGACGGCGTCTG

## (2) INFORMATION FOR SEQ. ID. NO.: 46:

### 40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

45

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear



(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

5 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....48

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:46:

10 GATCCGACGCCGTCGACCTCGTCGCCGACGCTGCCGACAGCGCCGACG

2) INFORMATION FOR SEQ. ID. NO.47: pcDNA3-1XUb-DEVD-Bla

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6459 base pairs

20 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

25

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

30 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6459

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:47:

35 gacggatcgggagatctccgatccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctgctgagtagtgcgcgagcaaaatctaagctacaacaaggcgaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcttgacattgattattga  
ctagtattaatagtaatacattacggggcattagttcatagcccatataggagtccgcgttacataacttacggtaaatggc  
ccgcttggtgaccgcccacgacccccgccattgacgtcaataatgacgtatgtcccatagtaacgccaatagggact  
40 ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagcatgacctatgggactttcctacttggcagtag  
atctacgtattagtcacgctattaccatgggtgatcggttttggcagtagcatcaatgggcgtggatagcggtttgactcacgg  
ggatttcaagtgctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
45 gaaccactgcttactggcttatcgaaattaatacgaactactataggagaccaagcttggtaccaccatggagatcttcg  
tgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgaccaa  
gacaagggaaggcatccctcctgaccagcagaggttgatcttctgctgggaacagctggaagatggacgcaccctgtctga  
ctacaacatccgaaagagtgccacctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcg



gtccaacccggtaagacacgacttatgccactggcagcagccactggtaacaggattagcagagcgaggtagtaggc  
 ggtgctacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatttggtagctgcgctctgctgaagcc  
 agttaccttcggaaaaagagttggtagctcttgatccggcaaaaccaccgctggtagcgggtggttttttggtaagca  
 gcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacggggctgacgctcagtggaaacgaaac  
 5 tcacgttaagggattttggtcatgagattatcaaaaaaggatcttcacctagatccttttaattaaaaatgaagtttaaatcaatc  
 taaagtatatatgagtaaacttggctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttc  
 atccatagttgcctgactccccgtcgtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatgata  
 ccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggggccgagcgcagaagtggg  
 cctgcaactttatccgctccatccagcttataattgttggcgggaagctagagtaagtagttcggcagtaatagtttgcgca  
 10 acgttgttgccattgctacaggcatcgtgggtgcacgctcgtcgtttggtaggttcattcagctccggttcccaacgatcaa  
 ggcgagttacatgatccccatgttgtgcaaaaaagcggtagctccttcggctcctccgatcgtgtgcagaagtaagttggcc  
 gcagtggtatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtga  
 gtactcaaccaagtcatctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggataataccgcg  
 ccacatagcagaactttaaaagtgtcatcattgaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgag  
 15 atccagttcgtatgaaccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaaca  
 ggaaggcaaaaatgccgaaaaaagggaataaggcgacacggaaatgtgaatactcactcttcttttcaatattattg  
 aagcatttatcaggggtattgtctcatgagcggatacatattgaatgtatttagaaaaataaacaataaggggtccgcgcaca  
 ttccccgaaaaagtgccacctgacgtc

20

(2) INFORMATION FOR SEQ. ID. NO:48:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 6726 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

35 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6726

40

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:48:

gacggatcgggagatctcccgatcccctatggtcactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
 ctccctgcttgtgtgttgagggtcgtgtagtgctgcgcagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
 gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcggttgacattgattattga  
 45 ctagtattaatagtaataattacgggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
 ccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
 ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccc

cctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagacacattatgggactttcctacttggcagtag  
 atctacgtattagtcacgctattaccatgggtgatcggtttggcagtagacatcaatgggcgtggatagcgggttgactcacgg  
 ggatttccaaagtctccacccattgacgtcaatgggagtttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
 5 gaaccactgcttactggcttatcgaatataacgactactatagggagaccaagcttgatcgaattctgcagccccg  
 ggggatctaccatggaaatcttcgtgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaagatccaagacaagggaagcatccctcctgaccagcagagggtgatcttctgctgggaaacagctg  
 gaagatggagcgaccctgtctgactacaacatccagaaagagtgccaccctgcacctggactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 10 agaatgtcaaggcaagatccaagacaagggaagcatccctcctgaccagcagagggtgatcttctgctgggaaacagctg  
 gaagatggagcgaccctgtctgactacaacatccagaaagagtgccaccctgcacctggactccgtctcagaggtgtgcac  
 cacggatccgtcggcgtctcggcagcgtcggcgacgaggtcgcggcgtcggatccggggcggtgctgcaccacaga  
 aacgctgggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatcacaacagcggtaa  
 gatccttgagagtttcccccgaagaacgtttccaatgatgagcacttttaagttctgctatgtggcgcggtattatccgt  
 15 attgacgccgggcaagagcaactcggtcggcgatacactattctcagaatgacttgggtgagtactaccagtcacagaa  
 aagcatcttacggatggcatgacagtaagagaattatgcagtgtcgcataaccatgagtataacactcggccaacttac  
 ttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttg  
 ggaaccggagctgaatgaagccatacacaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcg  
 caaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggagggcgataaagtgcagga  
 20 ccacttctgcgctcggcccttccggctgggtgtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcatt  
 gcagcactggggccagatggtaagccctccgtatcgtatgtatcagcagcggggagtcaggcaactatggatgaacg  
 aaatagacagatcgtgagataggtccctactgattaagcattgtaatctagaggggccctattctatagtgtcacctaaatg  
 ctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgtgtttgccccctccccgtgccttcttgacct  
 ggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagttaggtgtcattctattctgggg  
 25 ggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtgggggatgcgggtgggctctat  
 ggcttctgagggcgaaagaaccagctgggggtctagggggtatccccacgcgccctgtagcggcgcatgaagcgcggc  
 ggggtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccctagcggcgctccttctgctttcttcccttcttct  
 cgccacgttcggcgcttccctgtcaagctctaaatcggggcatcccttaggggttccgatttagtgccttacggcacctcga  
 ccccaaaaaacttgattaggtgatggttcacgtagtggccatcgcctgatagacggttttccgctttgacgttgaggtc  
 30 cacgttcttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggctattcttttgatttataagggattttg  
 gggatttcggcctattggttaaaaaatgagctgatttaacaaaaatgaacgcgaattaatctgtggaatgtgtgacgttagg  
 gtgtggaagtccccaggctccccaggcaggcagaagtatgaaagcatgcatctcaattagtcagcaaccaggtgtgga  
 aagccccaggctccccagcaggcagaagtatgaaagcatgcatctcaattagtcagcaaccatagtcgcccccactaact  
 ccgccatccccgccctaactccgccaggtccgccattctccgcccatggctgactaattttttattatgcagaggcc  
 35 gagggcgccctctgcctctgagctattccagaagtagtgaggaggctttttggaggccctaggcttttgcaaaaagctccgg  
 gagcttgatatccattttcgatctgatcaagagacaggatgaggatcgtttcgcagattgaacaagatggattgcacgca  
 ggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtg  
 ttccggctgtcagcgcaggggcgccgggttctttgtcaagaccacgttccgggtgccctgaatgaactgcaggacgag  
 gcagcgcggctatcgtggctggccacgacggcggttcttgcgagctgtgctcgcaggtgtcactgaagcgggaaggg  
 40 actggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggc  
 tgatgcaatgcggcggtgcatacgttgatccggctacctgcccattcgaccaccaagcgaacatcgcacgcagcgag  
 cacgtactcggatggaagccggcttctgtgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaact  
 gttcggcagggtcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgcgaatatca  
 tgggtggaanaatggccgctttctggattcatcactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggc  
 45 taccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacgggtatcgccgctcccgattcgc  
 agcgcacgccttctatgccttctgacgagttctctgagcgggactctgggggttgaaatgaccgaccaagcgacgccc  
 aacctgccaatcacgagatttcgattccaccgccgcttctatgaaaggttgggcttgggaatcgtttccgggacgccggctg

gatgatcctccagcgcggggatctcatgctggagttcttcccccacccaactgtttattgcagcttataatggttacaaataa  
 agcaatagcatcacaaatttcacaaataaagcattttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttacc  
 atgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatactgtttcctgtgtgaaattgttatccgctcaca  
 attccacacaacatacagccggaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactacattaattgcgt  
 5 tgcgctcactgcccgtttccagtcgggaaacctgtcgtccagctgcattaatgaatcggccaacgcgcggggagagggc  
 ggtttgcgtattggcgctcttccgcttctcgtcactgactcgtcgcgtcggctgttcggctgcggcgagcgggtatcagc  
 tctactcaaaggcggtaatacaggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaa  
 aaggccaggaaccgtaaaaaggccgcgttgcgtggcgtttttccataggctccgccccctgacgagcatcacaaaaatcg  
 acgctcaagtcagagggtggcgaaacccgacaggactataaagataccaggcggtttccccctggaagctccctcgtgcgt  
 10 ctctgttccgacctgcccgttaccggatcctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgct  
 gtaggtatctcagttcgggttaggtcgttcgtccaaagctgggctgtgtgcacgaacccccgttcagcccagccgtcgc  
 ccttatccggtaactatcgtcttgagccaacccggtgaagacacgacttatcgcactggcagcagccactggtaacaggat  
 tagcagagcgaggtatgtaggcgggtgtacagagtcttgaagtgggtggcctaactacgggtacactagaaggacagtatt  
 ggtatctgcgctctgtgaagccagttaccttcggaaaaagagttggtagctcttgatccggcacaacaaaccaccgctggt  
 15 gcggtggtttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatccttgatcttttctacggggt  
 ctgacgctcagtggaacgaaaaactcacgttaagggttttggctatgagattatcaaaaaggatcttcacctagatccttttaa  
 attaaaaatgaagtttttaaatcaatctaaagtatatatgagtaaaacttggtctgacagttaccaatgcttaatcagtgaggcacct  
 atctcagcgatctgtctatttcgttccatcatagttgcctgactccccgtcgtgtagataactacgatacgggaggggttaccat  
 ctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccagccgga  
 20 agggccgagcgcagaagtggctcctgcaactttatccgctccatccagcttattaattgttgcgggaagctagagtaagta  
 gttcggcagttaatgtttgcgcaacgttgttgcattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattc  
 agctccgggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggctccggt  
 cgttgcagaagtaagtggccgagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgta  
 agatgcttttctgtactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggc  
 25 gtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatttggaacggttcttcggggcgaaaactc  
 tcaaggatcttaccgctgttgagatccagttcgtatgaacccactcgtgcaccaactgatcttcagcatcttttactttacca  
 gcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatgttgaatact  
 catactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgtatttagaaaaata  
 aacaaataggggttccgcgcacatttccccgaaaagtccacactgacgtc  
 30

## (2) INFORMATION FOR SEQ. ID. NO:49:

35

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6969 base pairs

40

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

45

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5 (B) LOCATION: 1.....6969

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:49:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
10 gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagtatttaataagtaataacacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgcccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
15 atctacgtatttagtcgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaatcaacgggactttccaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggtaactaga  
gaacccactgcttactggcttaccgaaattatagcactcactatagggagaccaagcttgatcgaattcctgcagccccg  
ggggatctaccatggaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
20 agaattgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcacccctgcacctggtagctccgtctcagaggtgtgcac  
cacggatctaccatggaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
agaattgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcacccctgcacctggtagctccgtctcagaggtgtgcac  
25 cacggatctaccatggaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
agaattgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcacccctgcacctggtagctccgtctcagaggtgtgcac  
cacggatccgtcggcgctgtcggcagcgtcggcgacgaggtgcacggcggtcgatccggggcggtgctgcacccaga  
aacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggtttacatcgaactggatctcaacagcggtaa  
30 gatccttgagagtttcccccgaagaacgtttccaatgatgagcacttttaaaagtctgctatgtggcgcggtattatcccgt  
attgacggcgaggcaagagcaactcggctgccgcatacactattctcagaatgacttggtagtactaccagtcacagaa  
aagcatcttacggatggcatgacagtaagagaattatcgagtgtgccataacctagtgataacactgcggccaacttac  
ttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttg  
ggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatgcaacaacgttgcg  
35 caaactattaaactggcgaactacttacttagcttcccggcaacaattaatagactggatggagggcgataaagttgcagga  
ccacttctgcgtcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcatt  
gcagcactggggccagatggttaagccctccgtagctgtagttatctacacgacggggagtcaggcaactatggatgaacg  
aaatagacagatcgtgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgacacctaattg  
ctagagctcgtgacagcctcactgtgccttctagttgccagccatctgttgttggccctccccgtgccttcttgacctt  
40 ggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggg  
ggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctat  
ggcttctgagggcgaaagaaccagctggggctctagggggtatccccacgcgcctgtagcggcgacattaagcgcggc  
gggtgtggtgttacgcgcagcgtgaccgctacacttggcagcgccttagcggcgctccttctgcttcttcccttcttct  
cgccacgttcgccggcttccccgtcaagctctaaatcggggcatcccttaggggtccgatttagtgcttacggcacctcga  
45 ccccaaaaaacttgattagggtgatggttcacgtatgggccatcgcctgatagacggttttgcctttgacgttgaggagtc  
cacgttcttaatagtgactcttgttccaaactggaacaacactcaacctatctcgggtctattctttgattataagggattttg  
gggatttcggcctattgggttaaaaaatgagctgatttaacaaaaattaacgcgaattaattctgtggaatgtgtgcagttagg

gtgtggaagtcctccaggtcccccagggcagggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgga  
 aagtcctccaggtcccccagggcagggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcctcccccctaac  
 ccgcccataccgcccctaactccgcccaggtccgcccattctccgcccataggctgactaattttttattatgcagagggc  
 gagggccgctctgcctctgagctattccagaagtagtgaggagggctttttgaggccctaggcttttgcaaaaagtcctccgg  
 5 gagcttgatataccattttcgatctgatcaagagacaggatgaggatcgtttcgcattgaaacagatggattgcacgca  
 ggttctccggcggcttgagggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatccggcgtg  
 ttccggctgtcagcgcagggggcggcggttcttttgcagaccgacctgtccgggtgccctgaatgaactgcaggacgag  
 gcagcggcggctatcgtggctggccacgacggcggttccttgcgcagctgtgctcgcaggtgtcactgaagcgggaaggg  
 actggctgctattgggcgaagtggcggggcaggatctcctgtcatctcaccttgcctcgcggagaaagtatccatcatggc  
 10 tgatgcaatgcggcggctgcatacgttgatccggctacctgcccattcgaccaccaagcgaacatcgcacgcagcag  
 cacgtactcggatggaagccggcttctgtcgcagcaggtgatctggacgaagagcagcaggggctcgcggcagccgaact  
 gttcggcaggctcaaggcgcgcattgcccgcaggcggcaggtatctgtcgtgacctatggcgatgctgcttgcggaatatca  
 tggtggaataatggcggcttttctggttcatcgcactgtggcgggctgggtgtggcggaccgctatcaggacatagcgttggc  
 taccggtgatattgctgaagagcttggcggcgaaatgggctgaccgcttctcgtgctttacgggtatcggcgtcccgattcgc  
 15 agcgcacgccttctatgccttctgacgagttcttgcagcgggactctgggggttcgaaatgaccgaccaagcgcagccc  
 aacctgccatcacgagatttcgattccaccggcgccttctatgaaagggtgggcttcggaatcgtttccgggacggcggtg  
 gatgatctccagcgcggggatctcatgctggagttcttcgcccaccccaactgtttattgcagcttataatggttacaataa  
 agcaatagcatcacaaatttcacaaataaagcatttttctactgcattctagtgtgtgttgcgaaatcatcaatgatcttctc  
 atgtctgtataccgtcgacctctagctagagcttggcgtaacatggtcatagctgtttcctgtgtgaaattgttatccgctcaca  
 20 attccacacacatacagagccggaagcataaagtgtaaagcctgggggtgcctaatgagtgcagtaactacattatgtcgt  
 tgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattatgaatcggccaacgcgcggggagaggc  
 gggttgcgtattggcgcttctccgcttctcgtcactgactcgtcgcgtcggcgttcgggtgcggcagcgggtatcagc  
 tcactcaaaggcggtaatcgggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaa  
 aaggccaggaaccgtaaaaaggcgcgttgcgtggcgttttccataggctccgccccctgacgagcatcacaaaaatcg  
 25 acgctcaagtcagaggtggcgaaccgcagaggactataagataaccaggcgtttccccctggaagctccctcgtgcgt  
 ctctgttccgacctgccgcttaccggatacctgtccgccttctcccttcgggaagcgtggcgtttctcaatgtcacgct  
 gtaggtatctcagttcgggtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcg  
 ccttatccggtaactatcgtcttgatccaacccggtgaagacacgactatcggcactggcagcagccactggttaacaggat  
 tagcagagcggaggtatgtagcggtgctacagagttctgaagtgggtggcctaactacggctacactagaaggacagtattt  
 30 ggtatctgcgtctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaacaaccaccgctggta  
 gcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggt  
 ctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttta  
 attaaaaatgaagtttaaatcaatcaaatgaatgatagtaaaacttggtctgacagttaccaatgcttaacagtgaggcacct  
 atctcagcgtatctgtctatttcgttcacatagattgcctgactccccgtcgtgtagataactacgatacgggagggcttaccat  
 35 ctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccgga  
 agggccgagcgcagaagtggcctgcaactttatccgctccatccagcttattaattgttgcgggaagctagagtaagta  
 gttcggcagttaatagtttgcgaacgttgttgcattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttattc  
 agtccgggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggtagctccttcggtcctccgat  
 cgttgcagaagtaagtggccgagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgta  
 40 agatgcttttctgtgactggtgagttactcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttccccggc  
 gtcaatacgggataataccgcgccacatagcagaactttaaagtgtcatcattggaaaacgttcttcggggcgaaaactc  
 tcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatcttttactttacca  
 gcgtttctgggtgagcaaaaacaggaaggcaaaatgccgaaaaaagggaataaggcgacacggaaatgttgaatact  
 catactcttcttttcaatattattgaagcattatcagggttattgtctcatgagcggatcatatttgaatgtatttagaaaaata  
 45 aacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:50:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 7212 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

15

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20 (B) LOCATION: 1.....7212

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:50:

gacggatcgggagatctccgatccctatggcgaactctcagtacaatctgctctgatgcccatagttaagccagtatctg  
ctccctgcttgtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
25 gcatgaagaatctgcttagggttaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcttgacattgattattga  
ctagtattaatagtaataattacgggggtcattagttcatagcccatataggagttccgcgttacataaacttacggtaaatggc  
ccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgccc  
cctattgacgtcaatgacggtaaatggccgctggcattatgccagtagcatgacattatgggactttcctacttggcagtag  
30 atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaacccactgcttactggcttatcgaataatacgaactcactataggagacccaagcttgatacgaattctcgcagccccg  
ggggatctaccatggaaatctcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtacaccattg  
35 agaatgtcaaggcaaaatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaaagagtcacacctgacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatctcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaaagagtcacacctgacctgggtactccgtctcagaggtgtgcac  
40 cacggatctaccatggaaatctcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaaagagtcacacctgacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatctcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaaagagtcacacctgacctgggtactccgtctcagaggtgtgcac  
45 cacggatccgtcggcgtgtcggcagcgtcggcgacgaggtgcacggcgtcggatccggggcgtggctgcaccaga  
aacgctgggtgaaagttaaagatgctgaagatcagttgggtgcacgagtggtgtacatgaactggatctcaacagcggttaa



gatccttgagagttttgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccg  
 attgacgccgggcaagagcaactcggcgcgcacatactattctcagaatgacttggtgagtactaccagtcacagaa  
 aagcatcttacggatggcatgacagtaagagaattatgcagtgtgccataaccatgagtataactcgcggccaacttac  
 ttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttg  
 5 ggaaccggagctgaatgaagccatacacaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacggttgcg  
 caaactattaactggcgaactacttactctagcttcccggaacaattaatagactggatggaggcggataaagtgcagga  
 ccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtagcgtgggtctcgcggtatcatt  
 gcagcactggggccagatgtaagccctccgctatcgtagtattctacacgacggggagtcaggcaactatggatgaacg  
 aaatagacagatcgtgagataggtgcctcactgattaagcatttgtaaatctagaggccctattctatagtgacacataatg  
 10 ctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgtgttggccctccccgtgccttcttgacct  
 ggaaggtgccactcccactgtccttcttaataaaatgaggaaattgcacgcattgtctgagttaggtgcattctattctgggg  
 ggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctat  
 ggcttctgaggcggaaagaaccagctggggctctagggggtatcccgccgcttagcggcgccattaagcgcggc  
 ggggtgtggtggttacgcgcagcgtgaccgtacacttgccagcgccttagcggccgctccttctgcttcttcccttcttct  
 15 cgccacgttcgccggcttcccgtaagctctaaatcggggcatcccttaggggtccgatttagtgccttacggcacctcga  
 ccccaaaaaacttgattagggtgatgggtcacgtagtgggccatcgccctgatagacgggttttgcctttagcgttggagtc  
 caggttcttaatagtggactctgttccaaactggaacaacactcaaccctatctcggctctattctttgattataagggtttt  
 gggatttcggcctattggttaaaaaatgagctgatttaacaaaaattaacgcgaattaattctgtggaatgtgtgctcagttagg  
 gtgtggaagtcccgaggctcccgaggcaggcagaagtatgcaaaagcatgcattctcaattagtcagcaaccagggttgga  
 20 aagtccccaggctcccgagcaggcagaagtatgcaagcatgcattctcaattagtcagcaaccatagtcggcccttaact  
 ccggccatcccgccctaaactccggccagtccggccattctccggccatggctgactaattttttattatgcagaggcc  
 gaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttgaggccctagggcttttcaaaaagctccgg  
 gagcttctatccatttccggtatgatcaagagacaggatgaggatcgttccgatgattgaacaagatggattgcacgca  
 ggttctccggcgcttgggtggagaggctattcggctatgactgggcacacagacaatcggtgctctgatgccggcggtg  
 25 ttccggctgtcagcgcaggggcgccgggttctttgtcaagaccgacctgtccgggtgccctgaatgaactgcaggacgag  
 gcagcgcgggtatcgtggctggccacgacgggcgttcttgcgcagctgtgctcagctgtgactgaagcgggaagg  
 actggctgctattggcggaagtgcggggcaggatctcctgtcatctcacctgtcctgccgagaaagtatccatcatggc  
 tgatgcaatgcggcggtgcacacgcttgatccggctacctgccattcgaccaccaagcgaacatcgatcgagcagag  
 cacgtactcggatggaagccggtctgtcgtacaggatgatctggacgaagagcatcaggggctcgcgccagccgaact  
 30 gttcgccaggctcaaggcgcgcacatcccgcggcgaggatctcgtcgtgacctatggcgatgcctgcttgcggaatatca  
 tgggtggaatggccgctttctggattcatcgtactgtggcggtgggtgtggcgaccgctatcaggacatagcgttggc  
 taccgtgatattgtgaagagcttggcggcgaatgggtgaccgcttctcgtgctttacgggtatcgccgctcccattcgc  
 agcgcacgccttctatcgcttcttgacgagttcttctgagcgggactctgggggttcgaaatgaccgaccaagcgacggcc  
 aacctgccatcacgagatttcgattccaccgccgcttctatgaaggttggggttcggaatcgtttccgggacggcggtg  
 35 gatgatcctccagcgcggggatctcatgctggagtcttcgccaccccaacttgttattgcagcttataatggttacaaataa  
 agcaatagcatcacaatttcacaaataaagcatttttctactgattctagtgtggtttgtccaaactcatcaatgtatcttct  
 atgtctgtataccgtcgacctctagctagagcttggcgtaatcatggctatagctgttctctgtgtgaaattgttatccgctcaca  
 attccacacaacatacagccgggaagcataaagtgtaaagcctgggggtgcctaatgagttagtaactcacatttaattgcgt  
 tgcgctcactgcccgtttccagtcgggaaacctgtcgtccagctgcattaatgaatcgccaacgcgcggggagaggc  
 40 ggtttgcgtattgggcgcttctccgcttctcgtcactgactcgtcgcgtcgttcggtcggcgagcgggtatcagc  
 tactcaaaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaa  
 aaggccagggaaccgtaaaaaggccggttctgtggcgttttccataggctccggcccccctgacgagcatcacaataatcg  
 acgtcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgttccccctggaagctccctcgtcgcgt  
 ctctgttccgacctgcccgttaccggataacctgtccgccttctcccttcgggaagcgtggcgcttctcaatgctcacgt  
 45 gttaggtatctcagttcgtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcg  
 ccttatccggtaactatcgtcttgagtcacaacccggtaagacacgacttatcgccactggcagcagccactggttaacaggat  
 tagcagagcggaggtatgtaggcggtgtacagagttctgaagtggtggcctaactacggctacactagaaggacagtattt

5 ggatatcgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaaccaccgctggta  
 gcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggt  
 ctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacntagatccttttaa  
 attaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttgggtctgacagttaccaatgcttaacagttagggcacct  
 atctcagcgatctgtctatttcgttcacccatagttgctgactccccgtcgtgtagataactacgatacgggaggggttaccat  
 ctggccccagtgtgcaatgataccgcgagaccacgctcaccgggtccagatttatcagcaataaaccagccagccgga  
 agggccgagcgcagaagtggctcctgcaactttatccgcctccatccagctctattaattgttgccgggaagctagagtaagta  
 gttcgccagttaatagtttgcgaacgttggccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattc  
 agctccgggtcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggctctccgat  
 10 cgttgcagaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgta  
 agatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatcggcgaccgagttgctcttggccggc  
 gtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatttgaaaacgttcttcggggcgaaaactc  
 tcaaggatcttaccgctgttgagatccagttcgtatgtaaccactcgtgcaccaactgatcttcagcatcttttactttacca  
 gcgttctgggtgagcaaaaacagggaaggcaaaatgccgaaaaaagggaataaggcgacacggaaatgtgaataact  
 15 catactcttcttttcaatattattgaagcatttatcagggtattgtctcatgagcggatacatattgaatgtatttagaaaaata  
 aacaaataggggtccgcgcacatttccccgaaaagtccacctgacgtc

20 (2) INFORMATION FOR SEQ. ID. NO:51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....48

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:51:

GATCCGTCGGCGCTGTCGGCAGCGTCGGCGACGAGGTCGCTGGCGTCG

(2) INFORMATION FOR SEQ. ID. NO:52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

5 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

15 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....48

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:52:

20 GATCCGACGCCAGCGACCTCGTCGCCGACGCTGCCGACAGCGCCGACG

(2) INFORMATION FOR SEQ. ID. NO:53:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6459 base pairs

(B) TYPE: nucleic acid

30

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

35 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

40

(B) LOCATION: 1.....6459

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:53:

45 gacggatcgggagatctcccgatccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgctgtgtgttgaggctgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcttgacattgattattga  
ctagttattaatagtaataattacggggcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc

ccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttccatagtaacgccaatagggact  
 ttccattgacgtcaatgggtgactatttacggtaaactgccacttggcagtacatcaagtgtatcatatgccaaagtacgcc  
 cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagacattatgggactttctacttggcagtag  
 atctacgtattagtcacgtattaccatgggtgatcggttttggcagtagcatcaatgggcgtggatagcgggttactcacgg  
 5 ggatttccaaagtctccacccattgacgtcaatgggagtttgggttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
 gaaccactgcttactggcttaccgaaattaatacgaactactatagggagaccaagcttggtagccaccatggagatcttcg  
 tgaagactctgactggtgaagaccatcactctgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgaccaa  
 gacaaggaaggcatccctcctgaccagcagaggtgatcttggcggaaacagctggaagatggacgcaccctgtctga  
 10 ctacaacatccagaaagagtcaccctgcacgtgtagctcgtctcagaggtgtgacaccacggatccgtcggcgctgtcg  
 gcagcgtcggcgacgaggtcgtggcggtcggatccggggcggtggctgcaccagaaacgctgggtgaaagtaaaagatg  
 ctgaagatcagttgggtgcacgagtggttaccatcgaactggatctcaacagcggtaagatccttgagagtttcccccga  
 agaagcttttccaatgatgagcacttttaagtctgtatgtggcggttattatccgtattgacggcgggcaagagcaact  
 cggtcggcgcatatactattctcagaatgacttgggtgagtagtaccagtcacagaaaagcatcttaccggatggcatgaca  
 15 gtaagagaattatgcagtgtgccataaccatgagtataacactgcggcaacttacttctgacaacgatcggaggaccg  
 aaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagcca  
 taccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaactattaactggcgaactactta  
 ctctagcttcccggaacaattaatagactggatggagggcgataaagttgcaggaccacttctgcgctcggcccttccgg  
 ctggctggtttattgtgataaacttggagccgggtgagcgtgggtctcgcgggtatcattgcagcactggggccagatggtaa  
 20 gcccctccgtatcgtagtattctacacgacggggagtcaggcaactatggatgaacgaatagacagatcgtgagatagg  
 tgcctcactgattaagcattggtaatctagagggccctattctatagtgacctaataatgctagagctcgtgatcagcctcga  
 ctgtgccttctagtgtccagccatctgtgttggccctccccgtgccttcttgaccctggaaagtgccactccactgtcct  
 ttctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtgggtggggcaggacagca  
 aggggggaggttgggaagacaatagcaggcatgctggggatgcgggtgggctctatggcttctgagggcgaaagaacca  
 25 gctgggggctctagggggatccccacgcgccctgtagcggcgcatlaagcgcggcggtgtggttgggttacgcgcagcgt  
 gaccgctacacttgcagcggccctagcggccgctccttctgcttcttcccttcttctgcccacgttcgccggttccccgt  
 caagctctaatacggggcatccctttaggggtccgatttagtgctttacggcacctcaccaccaaaaaacttgattagggtgat  
 gggttacgtagtgggccatcggcctgatagacgggttttgcctttgacgttggagtcacgttcttaatagtgacttctgtt  
 ccaaaactggaacaacactcaaccctatctcggctctattctttgattataagggttttggggatttcggcctatttggttaaaaa  
 30 atgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaagtccccaggctcccc  
 aggcaggcagaagtagtcaaaagcatgcattcaattagtcagcaaccaggtgtggaagtccccaggctccccagcagg  
 cagaagtagtcaaaagcatgcattcaattagtcagcaaccatagtcggcccttaactccggccatcccggcccttaactccg  
 cccagttccgcccattctccgccccatggctgactaattttttattatgcagaggccgagggcgctctgctctgagctat  
 tccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagcttgtatatccatttccggatctg  
 35 atcaagagacaggatgaggatcgttccgcatgattgaacaagatggattgcacgcaggttctccggcgcttgggtggaga  
 ggctattcggctatgactgggcacaacagacaatcggtcgtctgatgccggcggttccggctgtcagcgcaggggcgcc  
 ccggttcttttgcagaccgacctgtccgggtgcctgaatgaactgcaggacgaggcagcgcggctatcgtggctggcc  
 acgacggcggttcttgcgcagctgtgtcgcaggttgcactgaagcgggaaggaggtggtgctattgggcgaagtgcc  
 ggggcaggatctcctgtcatctcacttgcctcctggcgagaaagtatccatcatggctgatgcaatgcggcggtgcatac  
 40 cttagtcgggtacctgcccattcgaccaccaagcgaacatcgcacgcagcagcagctactcggatggaagccggtctt  
 gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcagc  
 ccgacggcgaggatctcgtgtgacccatggcgatgcctgcttgcgaatatcatggtggaatggccgcttttctggatt  
 catcgaactgtggcggtgggtgtggcgaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttgg  
 cggcgaatgggctgaccgcttctcgtgtttacgggtatcgccgctcccgattcgcagcgcacgccttctatgccttctg  
 45 acgagttcttctgagcgggactctgggggtcgaatgaccgaccaagcgcgccaacctgccatcacgagatttcgattc  
 caccggcgcttctatgaaaggttggggttcggaatcgtttccgggacggcggtggatgatcctccagcgcggggatct  
 catgctggagttcttgcgccacccaactgtttattgcagcttataatggttacaaataaagcaatagcatcaaaatttcaca

aataaagcattttttactgcattctagttgtggttgccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctag  
 ctgagcgttggcgtaaatcatggtcatagctgttccctgtgtgaaattgttatccgctcacaattccacacaacatacagagccgg  
 aagcataaagtgtaaagcctgggggtgcctaatagtgtgagctaaactcacattaattgcgttgcgctcactgcccgtttccagt  
 cgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagagggcggttgcgtattgggcgctcttcc  
 5 gcttctcgtcactgactcgtcgcgtcggctggtcggctcggcgagcgggtatcagctcactcaaaggcggtaatacgg  
 ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag  
 gccgcgttgcgtggcgttttccataggtccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga  
 aacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgtctcctgttccgacctgccgtta  
 ccggatacctgtccgcttttctccctcgggaagcgtggcgttttctcaatgctcacgctgtaggtatctcagttcgggtgtagg  
 10 tcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttga  
 gtccaacccggtaagacacgactatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggc  
 ggtgctacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatttggatctgcgctctgctgaagcc  
 agttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcgggtgtttttgttgaagca  
 gcagattacgcgcagaaaaaaaggatctcaagaagatccttgatctttctacggggtctgacgctcagtggaacgaaaac  
 15 tcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttaaatcaatc  
 taaagtatatatgagtaaacttggctgacagttaccaatgcttaatacgtgagggacctatctcagcgatctgtctatttcgttc  
 atccatagttgcctgactccccgtcgtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatgata  
 ccgcgagaccacgctcaccggctccagatttatcgaataaaccagccagccggaagggccgagcgcagaagtggg  
 cctgcaactttatccgctccatccagctctatattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgca  
 20 acgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccgggtcccaacgatcaa  
 ggcgagttacatgatccccatgttggtgcaaaaaagcggttagctccttcggctcctccgacgttgcagaagtaagttggcc  
 gcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtga  
 gtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggataataccgcg  
 ccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgag  
 25 atccagttcgatgaaccactcgtgcaccaactgatctcagcatctttactttaccagcgtttctgggtgagcaaaaaca  
 ggaaggcaaaaatgccgaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcttttcaatattattg  
 aagcatttatcagggttattgtctcatgagcggatacatatttgatgtatttagaaaaataaacaataaggggtccgcgcaca  
 ttccccgaaaagtgccacctgacgtc

30

## (2) INFORMATION FOR SEQ. ID. NO:54:

### 35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6726 base pairs

(B) TYPE: nucleic acid

40

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

### 45 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6726

5

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:54:

gacggatcgggagatctcccgatcccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgtggaggtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
10 ctagtattataatagtaataaattacgggggtcattagttcatagcccataatggagttccgcgttacataacttacggtaaattggc  
ccgcttggtgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggact  
ttccattgacgtcaatgggtggactatttacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgcc  
cctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactacgg  
15 ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggtaactaga  
gaaccactgcttactggcttatcgaaattaatcagctactataggagaccaagcttgatcgaattcctgcagcccg  
ggggatctaccatggaaattctcgtgaagactctgactgtaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaagcatccctctgaccagcagaggttgatctttgctgggaaacagctg  
20 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaattctcgtgaagactctgactgtaagaccatcactctcgaagtggagccgagtacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaagcatccctctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatccgtcggcgtctgcggcagcgtcggcgacgaggtcgtggcgtcggatccggggcgtggctgcaccagaa  
25 acgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatcgaacagcggtaag  
atccttgagagtttccgcccgaagaacgttttccaatgatgagcacttttaagtctgctatgtggcgcggtattatcccgtat  
tgacgccgggcaagagcaactcggctcggcgatcacactattctcagaatgacttgggtgagtactaccagtcacagaaaa  
gcattctacggatggcatgacagtaagagaattatgcagtgtccataaccatgagtataactgcggccaacttacttc  
tgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttggga  
30 accggagctgaatgaagccatacacaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaa  
ctattaactggcgaactacttactctagcttccggcaacaattaatagactggatggagggcgataaagttgcaggaccact  
tctgcgctcggcccttccggctggctgtttattgctgataaatctggagccgggtgagcgtgggtctcgcggtatcattgcag  
cactggggccagatggtgaagccctcccgtatcgtattatctacacgacggggagtcaggcaactatggatgaacgaata  
gacagatcgtgagataggtgcctcactgattaagcattggtaatctagaggccctattctatagtgtcacctaatgctaga  
35 gctcgtgatcagcctcactgtgccttctagtgtccagccatctgttgtttgccctccccgtgccttcttgaccctggaa  
ggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtg  
gggtggggcgaggacgaagggggaggattgggaagacaatagcaggcatgctggggatcggtgggctctatggctt  
ctgaggcggaagaaccagctggggctctaggggtatccccacgcgcctgtagcggcgccattaagcggcggggtg  
tggtggttacgcgcagcgtgaccgtacacttgcagcgccctagcggcgctccttctccttctccttctcgcga  
40 cgttcgccggcttccccgtcaagctctaataatcgggcatccctttagggttccgatttagtgccttacggcacctcgacccca  
aaaaacttgattagggtgatggttcacgtagtggccatcggcctgatagacggttttcgcccttgacgttgaggtccacgt  
tctttaatagtggactctgttccaaactggaacaactcaaccctatctcggctctattcttttgattataagggaatttggggat  
ttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgcagttagggtgtgg  
aaagtccccagggtccccaggcaggcagaagtatgcaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcc  
45 ccaggctccccagcaggcagaagtatgcaagcatgcatctcaattagtcagcaaccatagtcgcccccactccgcc  
catcccgccctaactccgccagttccgccattctccgcccatggctgactaattttttattatgcagaggccgagggc  
cgctctgcctctgagctattccagaagtgtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagctt

gtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcattgaaacaagatggattgcacgcaggttctc  
 cggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgcccggtgttccgg  
 ctgtcagcgcagggggcgccgggttcttttgcgaagaccgacctgtccgggtgccctgaatgaactgcaggacgaggcagc  
 gcggctatcgtggctggccacgacggcggttccttgcgcagctgtgctgcaggttgcactgaagcgggaagggactggc  
 5 tgctattgggcgaagtgcggggcaggatctcctgtcatctcaccttgcctcgcgagaaagtatccatcatggctgatgc  
 aatgcggcggtgtcatacgcttgatccggctacctgccattcgaccaccaagcgaaacatcgcatcagcgcagcagcgtat  
 ctccgatggaagccggcttctgtcgcacaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcc  
 aggtcaaggcgcgcacgcccgcggcgaggatctcgtcgtgacctgagcgcgatgcctgcttgcgaatatcatggtgga  
 aaatggccgcttttctgattcatcgcactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccgt  
 10 gatattgctgaagagcttggcggcgaatgggctgaccttctcctgtgctttacgggtatcgccgctcccgttcgcagcgca  
 tcgcttctatcgcttcttgacgagttctctgagcgggactctggggttcgaaatgaccgaccaagcgacgccaacctg  
 ccatcacgagatttcgattccaccgccgcttctatgaaagggttgggcttcggaatcgtttccgggacgccggctggatgat  
 cctccagcgcggggatctcatgctggagttcttcgccaccccaacttgtttattgcagcttataatggttacaaataaagcaa  
 tagcatcacaaatttcacaaataaagcatttttctactgcattctagtgtgttgcctaaactcatcaatgtatcttcatgtct  
 15 gtataccgtcgacctctagctagagcttggcgtaatcatggctatagctgttccctgtgtgaaattgttaccgtcacaaatcca  
 cacaacatacagccggaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactcacattaattgcgttgcgct  
 cactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgcccaacgcgcggggagaggcggttgc  
 cgtattggcgctcttccgcttctcgtcactgactcgtcgcctcggctgttcggctgcggcgagcgggtatcagctcactc  
 aaaggcggtataacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggc  
 20 caggaaccgtaaaaaggccgcgttgcgtggcgttttccataggtccgccccctgacgagcaccacaaaaatcgacgtc  
 aagtcagaggtggcgaaacccgacaggactataaagatacaggcggttccccctggaagctccctcgtgcgtctcctgt  
 tccgacctgcccgttaccggatctgtccgcttctcccttcgggaagcgtggcgttctcaatgctcacgctgtaggtat  
 tctcagttcgggtgtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagccccaccgctgcgcttatcc  
 ggtaactatcgtcttgagccaacccggtaagacacgactatcgccactggcagcagccactggtaacaggattagcaga  
 25 gcgaggtatgtagcggtgtacagagttctgaagtgggtggcctaactacggctacactagaaggacagtatttggtatct  
 gcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcgggtg  
 gttttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatcttttctacggggtctgacgc  
 tcagtggaaacgaaaactcacgttaagggttttggctatgagattatcaaaaaggatcttcacctagatccttttaattaaaaa  
 tgaagtttaataatcaatcaaaagtatatatgagtaaaacttggctgtgacagttaccaatgcttaatcagtgaggcacctatctcagc  
 30 gatctgtctattctgttcacatagttgcctgactccccgtcgtgtagataactacgatacgggaggggttaccatctggccc  
 cagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccggaaggggcc  
 gagcgcagaagtggctcctgcaactttatccgctccatccagctctattaattgttccgggaagctagagtaagtagttcgcc  
 agttaatagtttgcgaacgttgttgcattgctacaggcatcgtggtgtcacgctcgtcgttggtaggttcattcagctcc  
 gggtcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgc  
 35 cagaagtaagttggccgcagtggtatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatg  
 cttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttgcggcggtcaat  
 acgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaag  
 gatcttaccgctgttgatccagttcgtatgaacccactcgtgacccaactgatcttcagcatctttactttaccagcggtt  
 40 ctgggtgagcaaaaacaggaaggcaaatgccgcaaaaaaggggaataaggcgacacggaaatgttgaatactcatact  
 ctctcttttcaatattattgaagcatttatcagggttattgtctatgagcggatacatatttgatgtattagaaaaataacaaa  
 taggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6969 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular
- (ii) MOLECULE TYPE: cDNA
- 10 (ix) FEATURE:
- (A) NAME / KEY: Coding Sequence
- 15 (B) LOCATION: 1.....6969

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:55:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgagggtcgtgtagtgctgcgcagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
20 gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcttgacattgattattga  
ctagttattaatagtaataattacggggctcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
ccgctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
25 atctacgtattagtcacgctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaacccactgcttactggcttattcgaataatacgaactcactataggagaccaagcttgatcgaattcctgcagccccg  
ggggatctaccatggaaatcttctgtaagactctgactggttaagaccatcactctcgaagtggagccgagtgacaccattg  
30 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
cacggatctaccatggaatcttctgtaagactctgactggttaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
35 cacggatctaccatggaatcttctgtaagactctgactggttaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
cacggatccgtcggcgtgtcggcagcgtcggcgacgaggtcgtggcgtcggatccggggcgtggctgcaccagaa  
acgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggttaag  
40 atccttgagagttttcggcccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtat  
tgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttgggtgagtactaccagtcacagaaaa  
gcattctacggatggcatgacagtaagagaattatgcagtgtccataacatgagtataactgcggccaacttacttc  
tgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttgga  
accggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaa  
45 ctattaactggcgaactacttactctagcttcccggaacaattaatagactggatggagggcgataaagttgcaggaccact  
tctgcgtcggcccttcgggtgggtgtttattgctgataaacttgagccggtgagcgtgggtctcgcggtatcattgcag  
cactggggccagatggtaagccctcccgtatcgtagtattctacacgacggggagtcaggcaactatggatgaacgaata



gacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgacctaataatgctaga  
 gctcgctgatcagcctcgactgtgccttctagtggccagccatctgttgttggccctccccgtgccttccttgacctggaa  
 ggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgacattctattctggggggtg  
 ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggctt  
 5 ctgagggcgaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcggtg  
 tgggtggttacgcgcagcgtgaccgctacacttgccagcggcctagcggcgcctcttcgcttcttcccttcttctgccca  
 cgttcgccgggtttccccgtcaagctctaaatcggggcatcccttaggggtccgatttagtgccttacggcacctcgaccca  
 aaaaacttgattaggggtgatggttcacgtatggggccatcgccctgatagacgggttttcgcccttgacgttgagtcacgt  
 tctttaatagtgactcttgttccaaactggaacaacactcaaccctatctcggctattcttttgattataagggttttggggat  
 10 ttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgcagttagggtgtgg  
 aaagtccccaggtccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaagtc  
 ccaggtccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggcccttaactccgcc  
 catccccgcccttaactccggccaggtccgccattctccggccctatggctgactaattttttattatgagaggccgaggc  
 cgctctgcctctgagctattccagaagttagtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagctt  
 15 gtataccattttcgatctgatcaagagacaggatgaggatcggttcgatgattgaacaagatggattgcacgcaggttctc  
 cggccgcttgggtggagaggctattcgctatgactgggcacaacagacaatcggctgctctgatccggcgtgttccgg  
 ctgtcagcgcagggggcgccgggttctttgtcaagaccgacctgtccgggtgccctgaatgaactgcaggacgaggcagc  
 gcggctatcggtgtggccacgacgggcgttccttgcgcagctgtgctcgacgttgactgaagcgggaagggactggc  
 tgctattgggcgaagtgcggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgc  
 20 aatgcggcggtgtcatagcttgatccggctacctgccattcgaccaccaagcgaacatcgatcgagcgagcacgta  
 ctcggtggaagccggtctgtcgtatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcc  
 aggtcaaggcgcgcagtcggcgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtgga  
 aaatggccgcttttctggattatcgactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggtaccgt  
 gatattgtgaagagcttggcgccgaatgggctgaccgcttctcgtgctttacgggtatcgccgctcccgattcgacgcga  
 25 tcgcttctatcgcttcttgacgagttctctgagcgggactctgggggtcgaaatgaccgaccaagcgacgcccacctg  
 ccatcacgagatttcgattccaccggcgcttctatgaaaggttgggcttcggaatcgtttccgggacgccgggtggtgat  
 cctccagcgcggggatctcatgctggagtcttcgccaccccaacttggttattgcagcttataatggttacaataaagcaa  
 tagcatcacaaatttcacaaataaagcatttttctactgcattctagtgtgttgcctaaactcatcaatgtatcttcatgtct  
 30 gtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgaaattgttatccgtcacaaftcca  
 cacaacatacagccggaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactcacattaattgcgttgcgct  
 cactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttg  
 cgtattggcgctcttccgcttctcgtcactgactcgtcgtcgctcggtcgttcgggtcgggcgagcggatcagctcactc  
 aaaggcggtataacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggc  
 35 caggaaccgtaaaaaggccgcgttgcgtggcgttttccataggctccggccctgacgagcatcacaaaaatcgacgtc  
 aagtcagaggtggcgaaaccgacaggactataaagataccaggcggttccccctggaagctccctcgtgcgtctcctgt  
 tccgacctgccgcttaccggatactgtccgcttctcccttcgggaagcgtggcgcttctcaatgctcacgctgtaggtat  
 tctcagttcgggtgtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttatcc  
 ggtaactatcgtcttgagccaacccggtgaagacacgactatcgccactggcagcagccactggtaacaggattagcaga  
 40 gcgaggtatgtaggggtgtacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatttggtatct  
 gcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaaccaccgctggtagcggtg  
 gttttttgttgcaagcagcagattacgcgcagaaaaaaggatcagaagatccttgatctttctacggggtctgacgc  
 tcagtgaacgaaaactcacgttaagggttttggatgagattatcaaaaaggatcttcacctagatccttttaattaaaaa  
 tgaagtttaataatcaatcaaatatagtaaaccttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagc  
 45 gatctgtctatttcgttcatcatagttgcctgactccccgtcgtgtagataactacgatacgggaggggttaccatctggccc  
 cagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccggaaggggc  
 gagcgcagaagtggctcgtcaactttatccgctccatccagctctattaattgttggcggaagctagagtaagtagttcgcc  
 agttaatagtttgcgaacgttggccattgctacaggcatcgtggtgtcacgctcgtcgttggatggcttcattcagctcc

ggttcccaacgatcaaggcgagttacatgatccccatggttgcaaaaaagcggttagctccttcggctcctccgatcgttgt  
cagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatg  
ctttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtaaat  
acgggataataccgcgccacatagcagaactttaaaagtgtctatcattggaaaacgttctcggggcgaaaactctcaag  
5 gatcttaccgctgttgagatccagttcgtatgaaccactcgtgcacccaactgatcttcagcatcttttactttaccagcgtt  
ctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaagggaataaggggcgacacggaaatgtgaatactcact  
cttctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgtatttagaaaaataacaaa  
taggggttccgcgcacatttccccgaaaagtccacctgacgtc

10

## (2) INFORMATION FOR SEQ. ID. NO:56:

### 15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7212 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

### 25 (ii) MOLECULE TYPE: cDNA

#### (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

30

(B) LOCATION: 1.....7212

### (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:56:

gacggatcgggagatctccgatcccctatggtcgactctcagtacaatctgctctgatgccgcataagtaagccagtatctg  
35 ctccctgcttgtgttggaggtcgctgagtgtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagtattaatagtaataattacggggcttagttcatagcccataatggagtccgcgttacataacttacggtaaattggc  
ccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtaacatcaagtgtatcatatgccaaagtacgccc  
40 cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtaacatgacattatgggactttcctacttggcagtac  
atctacgtattagtcacgtattaccatggtgatcggttttggcagtaacatcaatgggcgtggatagcggtttgactcacgg  
ggatttcaagctccacccattgacgtcaatgggagtttggcaccacaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaaccactgcttactggcttatcgaattaatcagactcactataggagaccaagcttgatcgaattcctgcagcccg  
45 ggggatctaccatggaaatctcgtgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaagatccaagacaagggaagcgtccctctgaccagcagaggttgatcttctgggaaacagctg  
gaagatggacgcacctgtctgactacaacatccagaaagagtcaccctgcacctggtactccgtctcagaggtgtgcac

cacggatctacatggaaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagagggtgtgcac  
 cacggatctacatggaaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 5 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagagggtgtgcac  
 cacggatctacatggaaatcttctgtaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaaggaaggcatccctctgaccagcagagggtgatctttgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctgggtactccgtctcagagggtgtgcac  
 10 cacggatccgtcggcgtctgtcggcagcgtcggcgacgaggtcgtcggcgatccggggcgtggctgcaccagaa  
 acgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggtaag  
 atccttgagagttttgccccgaagaacgtttccaatgatgagcacttttaaagtctgctatgtggcgcgggtattatcccgtat  
 tgacgccgggcaagagcaactcggcgcgcatacactattctcagaatgacttggttgagtactaccagtcacagaaaa  
 gcatcttacggatggcatgacagtaagagaattatgcagtctgccataacctgagtgataacactcgggccaacttacttc  
 15 tgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttggga  
 accggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaa  
 ctattaactggcgaactacttactctagcttccggcaacaattaatagactggatggagggcgataaagttgcaggaccact  
 tctgcgtcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcgggtatcattgcag  
 cactggggccagatggtgaagccctcccgtatcgtagtattctacacgacggggagtcaggcaactatggatgaacgaaata  
 20 gacagatcgtctagataggtgcctcactgattaagcattgtaacttagagggccctattctatagtgcacctaagtctaga  
 gctcgtctgacgcctcactgtgcttctagtgtccagccatctgttgtttgccccctccccgtgccttccctgaccctggaa  
 ggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtg  
 ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatggctt  
 ctgagggcgaaagaaccagctggggctctagggggtatccccacgcgcctgtagcggcgcatlaagcgcggcgggtg  
 25 tgggtgttacgcgcagcgtgaccgctacacttgccagcgccttagcggcgctccttctccttctccttctctcgcga  
 cgctcggcggttccccgtcaagcttaaatcggggcatcccttaggggtccgatttagtgccttacggcacctcgacccca  
 aaaaacttgattagggtgatggttcacgtatggggccatcgccctgatagacggttttgcctttagcgttgaggatccacgt  
 tctttaatagtgactctgttccaaactggaacaacactcaaccctatctcggctctattcttttgattataagggttttggggat  
 ttcggcctattggttaaaaaatgagctgatttaacaaaaatlaacgcgaattaattctgtggaatgtgtgcagtaggggtgtgg  
 30 aaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcactcaattagtcagcaaccatagtcggccctaacccgcc  
 ccaggctccccagcaggcagaagtatgcaaagcatgcactcaattagtcagcaaccatagtcggccctaacccgcc  
 catcccgccctaacccgccagttccgccattctccggcccatggctgactaattttttattatgcagaggccgaggc  
 cgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagctt  
 gtatatccatttccgatctgatcaagagacaggatgaggatcgttctgcagattgaacaagatggattgcacgcaggttctc  
 35 cggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatccgccgtgttccgg  
 ctgtcagcgcaggggcgccccggttctttgtcaagaccgacctgtccggtgcctgaatgaactgcaggacgaggcagc  
 gcggctatcgtggctggccacgacgggcgttcttgcgcagctgtgctcgcaggttgactgaagcgggaaggactggc  
 tgctattgggcgaagtgcggggcaggatctcctgtcatctcacctgtcctcggagaaagtatccatcatggctgatgc  
 aatcgggcggctgcatacgttgatccggctacctgccattcgaccaccaagcgaaacatcgcacgagcgagcacgta  
 40 ctccgatggaagccggtcttgcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcc  
 aggctcaaggcgcgcgatcccgacggcgaggatctcgtcgtgacctatggcgatgcctgcttccgaatatcatggtgga  
 aaatggccgcttttctggattcatcagctgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccgt  
 gatattgctgaagagcttggcggcgaatgggctgaccgcttctcgtgctttacgggtatcggcgctcccgatcgcagcgca  
 tcgcttctatcgccttctgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctg  
 45 ccatcacgagatttcgattccaccgcccttctatgaaaggttgggcttcggaatcgtttccgggacgccggctggatgat  
 cctccagcgcggggatctcatgctggagttcttcgcccacccaactgtttattgcagcttataatggttacaaataaagcaa  
 tagcatcacaatttcacaaataaagcatttttctactgcattctagtgtgttgccttcaactcatcaatgatcttatcatgtct

gtataccgtcgacctctagctagagcttggcgtaatcatggcatagctgtttcctgtgtgaaattgttatccgctcacaaattcca  
cacaacatacagagccggaagcataaagtgtaaagcctggggtgcctaagagttagctaaactcacattaattgcgttgcgct  
cactgcccgcgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttg  
cgtattgggcgctcttccgcttctcgtcactgactcgtcgcgtcggcgttcggctgcggcgagcggtatcagctcactc  
5 aaaggcggtataacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggc  
caggaaccgtaaaaaggccgcgttgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctc  
aagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtcgcgtctcctgt  
tccgaccctgccgcttaccggatacctgtccgcttttccctcgggaagcgtggcgcttttcaatgctcacgctgtaggta  
tctcagttcgggtgtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgccttatcc  
10 ggtaactatcgtcttgagtccaacccgtaagacacgacttatcgccactggcagcagccactggtaacaggattagcaga  
gcgaggtatgtaggcggtgtacagagttcttgaaagtgtggcctaactacggctacactagaaggacagtatttggatct  
gcgctctgtgaagccagttaccttcggaaaaagagttgtagctcttgatccggcaacaaaccaccgctggtagcgggtg  
gttttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatcttttctacggggctcgacgc  
tcagtggaaacgaaaactcacgttaagggttttggatgagattatcaaaaaggatcttcacctagatccttttaattaaaaa  
15 tgaagttttaaatcaatctaagtatatatgagtaaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagc  
gatctgtctatttcgttcacatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccc  
cagtgtgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggcc  
gagcgcagaagtgtcctgcaactttatccgctccatccagttatattgttgccgggaagctagagtaagtagttcgc  
agttaatagtttgcgaacgttggtgccattgctacaggcctcgtggtgtcacgctcgtcgtttggtatggcttcattcagctcc  
20 ggttcccaacgatcaaggcgagttacatgatccccatgttggtcaaaaaagcggttagctccttcggctcctccgatcgttgt  
cagaagtaagttggccgagtggtatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatg  
cttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtaaat  
acgggataataccgcgccacatagcagaactttaaagtgtcgtcattggaaaacgttctcggggcgaaaactctcaag  
gatcttaccgctgttgagatccagttcgtgtaaccactcgtgcaccaactgatcttcagcatcttttactttaccagcggtt  
25 ctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataaggcgacacggaaatgttgaaactcatact  
cttctctttcaatattattgaagcattatcagggtattgtctcatgagcggatacatatttgatgtatttagaaaaataacaaa  
taggggttccgcgcacatttccccgaaaagtgccacctgacgtc

## 30 2) INFORMATION FOR SEQ. ID. NO.57: rhinovirus 14 2a

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1095 base pairs

35

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

45

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....1095

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:57:

5 ttgggtcgtgcagcttgtgtgcatgtaactgaaatacaaaacaaagatgctactggaatagataatcacagagaagcaaaatt  
gttcaatgattggaaaatcaacctgtccagcctgtccaacttagaaagaaactggaactcttcaacttatgttaggtttgattctg  
agtataccatactggccactgcatctcaacctgattcagcaaaactatcaagcaatttggtggtccaagccatgtatgttccac  
atggtgccccgaaatccaaaagagtgggcgattacacatggcaaagtgttcaaacccagtgattcttcaaggtggggg  
atacatcaaggtttagtgtgccttatgtaggattggcatcagcatataattgttttatgatggttactcacatgatgatgcagaaa  
ctcagtatggcataactgttctaaccatatgggtagtatggcattcagaatagtaaataaacatgatgaacacaaaactcttg  
10 tcaagatcagagtttatcacaggcgaaagctcgtgaagcatggattccaagagcaccagagcactaccctacacatcaa  
tagggcgacaaaattatcctaagaatacagaaccagtaattaagaagaggaaaggtgacattaaatcctatggtttaggacc  
taggtacggtgggatttatacatcaaatgttaaaataatgaattaccacttgatgacaccagaagaccaccataatctgatagc  
accctatccaaatagagatttagcaatagtctcaacaggaggacatggtgcagaaacaataccacactgaaccgtacatca  
ggtgtttactattccacatattacagaaagtattacccataatttgcgaaaagcccaccaacatctggattgaaggaagccct  
15 tattacccaagtagatttcaagcaggagtgtatgaaaggggttgggccgagagctaggagactgcggtgggattttgag  
atgcatacatgggccattggattgtaacagctgaaggtagtggatatgtttgtttgctgacatacgacagttggagtgtatc  
gcagaggaacag

2) INFORMATION FOR SEQ. ID. NO.58: HRV145' primer

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

25 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

35 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:58:

40 taggatccttgggtcgtgcagcttgtgtg

(2) INFORMATION FOR SEQ. ID. NO:59:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 nucleotides

(B) TYPE: nucleic acid

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

15 (B) LOCATION: 1.....29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:59:  
aaggatccctgttcctctgccatacactc

20

(2) INFORMATION FOR SEQ. ID. NO:60:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8022 base pairs

(B) TYPE: nucleic acid

30

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

35 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

40

(B) LOCATION: 1.....8022

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:60:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
45 ctccctgcttgtgttgagggtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaatagtaataatcacggggtcattagttcatagccatatatggagttccgcgttacataacttacggtaaattggc



ggaacaacactcaaccctatctcgggtctattcttttgatttataagggttttggggatttcggcctattgggttaaaaaatgagct  
 gatttaacaaaaatitaacgcgaattaattctgtggaatgtgtgtaggtgtggaaagtccccaggctccccaggcag  
 gcagaagtatgcaaagcatgcatctcaattagtcagcaaccagggtgtggaaagtccccaggctccccaggcagcagaagt  
 atgcaaagcatgcatctcaattagtcagcaaccatagtcggccctaaactccggccatcccggccctaaactccggccagtt  
 5 ccggccattctccggcccatggctgactaattttttattatgacagaggccgaggccgctctgctctgagctattccagaa  
 gtagtgaggaggctttttggaggccctaggcttttgcaaaaagctcccgggagcttgatatccatttcggatctgatcaaga  
 gacaggatgaggatcgttgcgatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctatt  
 cggctatgactgggcacaaacagacaatcggctgctctgatgccgccgtgtccggctgtcagcgcagggggcggccggctt  
 ttttgcagaccgacctgtccgggtgcctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacg  
 10 ggcgttcttgcgcagctgtgctcgcaggttgactgaagcgggaaggaggtggtgctattgggcgaagtggcggggca  
 ggatctctgtcatctacattgctcctgccgagaaagtatccatcatggtgatgcaatgcggcggctgcatacgttgatc  
 cggctacctgcccattcgaccaccaagcgaacatcgcacgcagcgcagcgtactcggatggaagccggcttctgctgat  
 caggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcattgcccgcac  
 ggcgaggatctcgtcgtgacctatggcgatgcctgcttgcgaatatcgttggaataaggcgcgttttctggattcatcga  
 15 ctgtggccggctgggtgtggcggaccgctatcaggacatagcgttggtacctccgtgatattgctgaagagcttggcggcg  
 aatgggctgaccgcttctcgtgctttacgggtatcggcgtcccgattcgcagcgcacgccttctatgccttctgacgagt  
 tcttctgagcgggactctggggttcgaaatgaccgaccaagcgcagcccaacctgccatcacgagatttcgattccaccgc  
 cgccttctatgaaagggtggggttcggaatcgtttccgggacggcggtggtgatcctccagcgcggggatctcatgctg  
 gagtcttcgccaccaccaactgtttattgcagcttataatggttacaaataaagcaatagcatcaaaatttcacaaataaag  
 20 cattttttactgcattctagtgtgtgttgcctaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagc  
 ttggcgtaatatggtcatagctgttctctgtgaaattgttatccgctcacaattccacacaacatacagccggaagcata  
 aagtgtaaagcctgggggtgcctaatgagtgcagtaactcacattaattgcgttgcgctcactgcccgcttccagtcgggaaa  
 cctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagagggcggttgcgtattgggcgcttccgcttctcgc  
 ctactgactcgtcgcctcggctgttcggctgcggcgagcgggtatcagctcactcaaggcggtatacggttatccaca  
 25 gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgtt  
 gctggcggttttccataggctccggcccccgtgacgagcatcacaaaaatcgcgctcaagtcagaggtggcgaaaccga  
 caggactataaagataaccaggcggttccccctggaagctccctcgtgcgctcctgttccgacctgcccgttaccggata  
 cctgtccgcttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcgggtgtaggtcgttcgc  
 tccaagctgggctgtgtgcacgaacccccgttcagcccagcgcgttaccggtaactatcgtcttgagtccaacc  
 30 cggttaagacacgactatcgcactggcagcagccactggtaacaggattagcagagcagggtatgtaggcgggtgctaca  
 gagtcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacctc  
 ggaaaaagagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtgtttttgttgcgaagcagcagattac  
 gcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaaacgaaaactcacgttaa  
 gggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtata  
 35 tatgagtaaaacttggtctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttcgttcacatagtt  
 gcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgtgcaatgataccgcgagac  
 ccacgctcaccggctccagattatcagcaataaaccagccagccgggaaggccgagcgcagaagtggctctgcaacttt  
 atccgctccatccagcttattaattgttggcgggaagctagagtaagtagttcggcagttaatagtttgcgcaacgttggtgcc  
 attgctacaggcctcgtggtgtcacgctcgtcgttgggtatgggttcattcagctccgggtcccaacgatcaaggcgagttaca  
 40 tgatcccccatgttgcacaaaaagcgggtagctccttcggctcctcgatcgttgcagaagtaagttggccgcagtggtatca  
 ctcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtactggtgagtactcaaccaa  
 gtcattctgagaatagtgtatgcggcgaccgagtgctcttggccggcgtcaatacgggataataccgcgccacatagcag  
 aactttaaaagtgtcatcattggaacacgttcttcggggcgaaaactcgaaggatcttaccgctgttgagatccagttcgat  
 gtaaccactcgtgcaccaactgatcttcagcatctttactttaccagcgtttctgggtgagcaaaaacaggaaggcaaaa  
 45 atgccgcaaaaagggaataaggcgacacggaaatgttgaatactcatactcttcttttcaatattgaagcatttatca  
 ggggtattgtctcatgagcggatacatattgaatgtatttagaaaaataacaaataggggttccgcgcacatttccccgaaa  
 agtgccacctgacgtc



2) INFORMATION FOR SEQ. ID. NO.61: rhinovirus 16 2a

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 636 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20 (B) LOCATION: 1.....636

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:61:

25 atgggaactttgtgtcgcgtattgtgaccagtgagcaattacacaaagtcaaagtggtacaaggatatcacaaagccaa  
acacaccaaagcttggtgcccaagaccacccagagctgttcaatactcacatacacataccaccaactacaaattgagttca  
gaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagtgcacatgtatgtcatgttggtaat  
ctaatacacagaaatctacatttattaactctgacatacatgattccattttagtgtcttattcatcagatttaacatataccgaac  
aagcacacaaggtgatggttatattccaacatgtaattgcactgaagctacatattactgcaaacacaaaaacaggtactacc  
caattaatgtcacacctcatgactggtatgagatacaagagagtgaatattatccaaaacatatccagtacaatttactaatag  
30 gtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgataggtattattacagcaggtggt  
gagggccatgttgcatcatagatcttagacactttcactgtgctgaa

2) INFORMATION FOR SEQ. ID. NO.62: HRV165 primer

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

40 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5 (B) LOCATION: 1.....29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:62:

aaggatccatgggaactttgtgttcgct

10

(2) INFORMATION FOR SEQ. ID. NO:63:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

25

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

30 (B) LOCATION: 1.....29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:63:

ttggatccttcttcagcacagtgaagtgc

35

(2) INFORMATION FOR SEQ. ID. NO:64:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7563 base pairs

(B) TYPE: nucleic acid

45

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

5 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7563

10

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:64:

gacggatcgggagatctcccgatcccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgtgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
15 ctagtatttaataagtaataacacggggtcattagttcatagcccatataggagtccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttccatagtaacgccaataggggact  
ttccattgacgtcaatgggtggactatttacggtaactgccacttggcagtacatcaagtgtatcatatgccaagtacggcc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtcacatgacctatgggactttctacttggcagtac  
atctacgtatttagtcacgctattaccatgggtgatgcggttttggcagtcacatcaatgggcgtggatagcgggttgactcacgg  
20 ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaacccactgcttactggcttctcgaatttaatacgaactcactatagggagaccaagcttgatatacgaattcctgcagcccc  
ggggatctaccatggaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaatgacaaaggaaggcattccctcctgaccagcagaggttgatctttgctgggaaacagctg  
25 gaagatggacgacccctgtctgactacaacatccagaaagagtcacccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaatgacaaaggaaggcattccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgacccctgtctgactacaacatccagaaagagtcacccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
30 agaatgtcaaggcaaatgacaaaggaaggcattccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgacccctgtctgactacaacatccagaaagagtcacccctgcacctgggtactccgtctcagaggtgtgcac  
cacggatccatgggaaactttgtgttcgcgtattgtgaccagtgaagcaattacacaaagtcaaaagtgttaacaaggatataca  
caaagccaaacacaccaaagcttggtgccccagaccaccagagctgttcaatactcacatacacataaccaccaactaca  
aattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagtgcacatgtatgtg  
35 catgttgtaataataatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaac  
atataccgaacaagcacacaagggtgatgggtatttccaacatgtaattgactgaagctacatattactgcaaacacaaaaa  
cagggtactaccaattaatgtcacacctcatgactggatgagatacaagagagtgatattatccaaaacatatccagtaca  
atttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgataggtattatta  
cagcaggtgggtgagggccatgttgattcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcaccc  
40 agaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggggttacatcgaactggatcacaacagcg  
gtaagatccttgagagttttgccccgaagaacgttttcaatgatgagcacttttaagttctgctatgtggcgcggtattatc  
ccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttgggtgagtactaccagtcaca  
gaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtgccataaccatgagtataacactgcggccaac  
ttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcg  
45 ttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaaacggttg  
cgaaactatttaactggcgaactacttacttagcttccggcaacaattaatagactggatggagggcgataaagtgcag  
gaccacttctgcgtcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatc

attgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa  
 cgaaatagacagatcgctgagataggtgctcactgattaagcattggtaatctagaggccctattctatagtgacactaa  
 atgctagagctcgctgatcagcctcgactgtgcttctagtgtccagccatctgtgtttgccccccccgtgcttccctgac  
 cctggaaaggtgccactcccactgtccttccataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctg  
 5 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggct  
 ctatggcttctgaggcggaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcaatgaagcg  
 ggcgggtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccctagcggcgctccttctgcttcttcccttct  
 ttctcgccacgttcgccggcttccccgtcaagctctaaatcggggcatcccttaggggtccgatttagtctttacggcacct  
 cgacccccaaaaaacttgattaggggtgatggttcacgtagtgggccatcgccctgatagacgggttttcgcccttgacgttgg  
 10 agtccacgttcttaatagtgactctgttccaaactggaacaacactcaaccctatctcggtctattctttgattataaggga  
 ttttggggatttcggcctattggtaaaaaatgagctgatttaacaaaatttaacgcgaattaattctgtggaatgtgtgcatt  
 aggggtgtggaagtccccagggtccccaggcaggcagaagtatgcaagcatgcattctcaattagtcagcaaccagggtgt  
 ggaaagtccccagggtccccagcaggcagaagtatgcaagcatgcattctcaattagtcagcaaccatagtcgccccct  
 aactccgccatccccgccctaactccgccagttccgccattctccgcccatggctgactaatttttttattatgcagag  
 15 gccgaggccgctctgctctgagctattccagaagttagggaggctttttggaggcctaggcttttgcaaaaagctcc  
 cgggagcttgatatccatttcggatctgatcaagagacaggatgaggatcgttccgatgattgaacaagatggattgcac  
 gcaggttctccggcgcttgggtggagaggctattcggtatgactgggcacaacagacaatcggctgcttgatccgc  
 cgtgttccggctgtcagcgcagggcgccccggttctttgtcaagaccgacctgtccggtgccctgaatgaactgcagga  
 cgaggcagcgcggctatcgtggctggccacgacggcgcttcttgcgcagctgtgctcgacgttgactgaagcggga  
 20 agggactggctgctattgggcgaagtggcgggcaggatctctgtcatctaccttgctcctgccgagaaagtatccatca  
 tggctgatgcaatgcggcggtcgcacagcttgatccggctacctgccattcgaccaccaagcgaaacatcgatcgag  
 cgagcacgtactcggatggaagccggtcttgcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccg  
 aactgttcgccaggctcaaggcgcgcagatcccgcagggcagggatctcgtcgtgacctggcgatgcctgcttccgaat  
 atcatgggtggaatggccgcttcttgattcagctgctggccggctgggtgtggcgaccgctatcaggacatagcgt  
 25 tggctaccgctgatattgtgaagagcttggcgggcgaatgggtgaccgcttctcgtgctttacgggtatcgccgtcccgat  
 tcgcagcgcacgccttctatcgcttcttgacgagttcttctgagcgggactctgggggtcgaatgaccgaccaagcgac  
 gcccacctgccatcacgagattcgattccaccgccgcttctatgaaggttgggcttcggaatcgtttccgggacgcc  
 ggctggatgatcctccagcgcggggatctcatgctggagtcttcgccaccccaacttgttattgcagcttataatggttac  
 aaataaagcaatagcatcacaatttcacaaataaagcatttttactgcattctagtgtgtgttgcacaaactcatcaatgta  
 30 tcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgttatccg  
 ctcaaatccacacaacatacagccggaaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactcacatta  
 attgcgttgcgctcactgcccgcttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcgggg  
 agaggcggttgcgtattggcgcttccgcttctcgtcactgactcgtcgcgtcggctgttcggctgcggcgagcgg  
 tatcagctcactcaaaggcggtataacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc  
 35 cagcaaaaggccaggaaccgtaaaaaggccggtgctggcggttttccataggtccgccccctgacgagcatcaca  
 aaatcgacgctcaagtcagaggtggcgaaaccgacaggactataaagataaccaggcggttccccctggaagctccctcg  
 tgcgtctcctgttccgacctgccgcttaccggatacctgtccgccttctcccttcgggaagcgtggcgcttctcaatgct  
 cacgctgtaggtatctcagttcgggtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagccgaccg  
 ctgcgcttatccggtaactatcgtcttgatccaaccggtaagacacgacttatcggcactggcagcagccactggtaac  
 40 aggtatagcagagcgagggtatgtaggggtgtacagagttctgaagtgggtggcctaactacgggtacactagaaggac  
 agtatttggatctgcgctctgctgaagccagttaccttcgaaaaagagttgtagctcttgatccggcaacaaaccaccg  
 ctggtagcgggtgtttttgttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctac  
 ggggtctgacgctcagtggaacgaaaactcacgttaagggaatttgggtcatgagattatcaaaaaggatcttcacctagatcc  
 ttttaaatataaataagatttaaatcaatctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaacagtgagg  
 45 cacctatctcagcgtatctgtctatttcgttatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggtt  
 accatctggccccagtgtgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccag  
 ccggaaggggcgcagcgcagaagtggctctgcaactttatccgctccatccagcttattaattgttccgggaagctagagt

aagtagttccagttaatagtttgcgcaacgttggtgccattgctacaggcatcgtggtgtcacgctcgtcgttggatggct  
tcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttggtgcaaaaaagcggttagctcctcgggtcc  
tccgatcgttggtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattcttactgtcatgccat  
ccgtaagatgcttttctgtgactggtgagttactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgc  
5 cggcgtcaatacgggataataccgcgccacatagcagaactttaaagtgtcatcattggaaaacgttctcggggcgaa  
aacttcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatctttactttc  
accagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataaggcgacacggaaatgttga  
atactcatactcttcttttcaatattattgaagcattatcagggttattgtctcatgagcggatacatatttgaatgtatttagaa  
aataaacaatatgggggtccgcgcacattccccgaaaagtgccacctgacgtc  
10

15 (2) INFORMATION FOR SEQ. ID. NO:65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7053 base pairs

20

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

25 (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

30

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7053

35 (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:65:

gacggatcgggagatctccgatccccatggtcgactctcagtacaatctgctctgatgcccatagtaagccagtatctg  
ctccctgcttgtgttgagggtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgttttgcgtcctcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagttattaatagtaaatcaattacggggcattagttcatagcccatataggagttccgcgttacataacttacggtaaaggc  
40 ccgcctgggtgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtgactatttacggtaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgcc  
cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagcatgaccttattgggactttcctacttggcagtag  
atctacgtattagtcacgtctattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaatcaacgggactttccaaatgtcgtaac  
45 aactccgccccattgacgcaaatggcggttagcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaaccactgcttactggcttatcgaaattaatcagactcactatagggagaccaagcttggtaccacatggagatcttcg  
tgaagactctgactggttagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgaccaa

gacaaggaaggcatccctcctgaccagcagaggtgatcttctgctgggaaacagctggaagatggacgcaccctgtctga  
ctacaacatccagaaaagagtgccaccctgcacctggtactccgtctcagaggtgggatgcacggatccatgggaactttgtgt  
tcgctgattgtgaccagtgcagcaattacacaaagtcacaaagtgtaacaaggatatcacaaagccaaacacaccaaaagctt  
5 ggtgcccagaccaccagagctgttcaatactcacatacacataccaccaactacaaattgagttcagaagtacacaatga  
tgtggctataagacctagaacaaactaacaactgttgggcctagtacatgtatgtcatgttggtatctaataacagaaat  
ctacatttatttaactctgacatacatgattccatttttagtgtcttattcatcagatttaatatataccgaacaagcacacaagggtg  
atggttatattccaacatgaattgcactgaagctacatattactgcaaacacaaaaacagggtactaccaattaatgtcacac  
ctcatgactggtatgagatacaagagagtgatattatccaaaacatccagtacaattactaataaggtgaaggaccatgtg  
aaccaggtgattgtggtgggaaattattatgcaaacatggagtgtatggtattattacagcaggtggtgagggccatgttgca  
10 ttcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcaccagaaacgctggtgaaagtaaaagatg  
ctgaagatcagttgggtgcacgagtggttacatcgaactggatcgaacagcggtaagatccttgagagtttgcggccga  
agaacgtttccaatgatgagcacttttaagtctgctatgtggcggtattatcccgtattgacgccgggcaagagcaact  
cggctgccgcatacactattctcagaatgacttgggtgagtactaccagtcacagaaaagcatcttaccggatggcatgaca  
gtaagagaattatgcagtgtgccataacctgagtataactgcggccaacttacttctgacaacgatcggaggaccg  
15 aaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagcca  
taccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaactattaactggcgaactactta  
cttagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccgg  
ctggtggtttattgtgataaacttgagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaa  
gcccctccgtatcgtagtattctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagatagg  
20 tgcctcactgattaagcattgtaacttagaggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctcga  
ctgtgccttctagtgtccagccatctgtgttggccctccccgtgccttcttgacctggaaggtgccactcccactgtcct  
ttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagca  
aggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggtctatggcttctgaggcggaaagaacca  
gctggggctctagggggtatccccacgcgcctgtagcggcgcaatgaagcgcggcggtgtggtgtgttacgcgcagcgt  
25 gaccgctacacttgccagcgccttagcgccttctccttctccttctccttctcgcacgttcgcggcttccccgt  
caagctctaatacggggcatcccttaggggtccgatttagtgccttacggcacctcgaacccaaaaaacttgattagggtgat  
ggttcacgtagtgggccatcgcctgatagacggttttgcgccttgacgttgagtcacgttctttaaagtggactctgtt  
ccaaactggaacaacactcaacctatctcggctattcttttgattataagggttttggggatttcggcctattggttaaaaa  
atgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaagtccccaggctcccc  
30 aggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaagtccccaggctccccagcagg  
cagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggcccttaactccgccatcccggcccttaactccg  
cccagttccggccattctccggccatggctgactaattttttatttatgcagaggccgaggccgctctgcctctgagctat  
tccagaagttagtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagcttgcataccatttcggatctg  
atcaagagacaggatgaggatcgttgcgtgattgaacaagatggattgcacgcaggttctccggccgcttgggtggaga  
35 ggctattcggctatgactgggcacaacagacaatcggtgctctgatgccgctgttccggctgtcagcgcaggggcgc  
ccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggcc  
acgacgggcgttcttgcgcagctgtgctcgcgtgtgactgaagcgggaagggactggctgctattgggcgaagtgcc  
ggggcaggatctcctgtcatctcaccttgcctgccgagaaagtatccatcgtgctgatgcaatgcggcggctgcatac  
cttgatccggctacctgccatcgaaccacaagcgaacatcgcacgagcagcagctactcggatggaagccggtctt  
40 gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcacatgc  
ccgacggcgaggatctcgtcgtgacctatggcgtgcttgcggaatcatggtggaatggccgctttctggtatt  
catcgaactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggctaccctgatattgctgaagagcttgg  
cggcgaaatgggctgaccgcttctcgtgctttacgggtatcgcgcgtcccgattcgcagcgcacgccttctatgccttctt  
acgagtttcttgagcgggactctggggttcgaaatgaccgaccaagcgcgccaacctgccatcacgagatttcgattc  
45 caccggcgcttctatgaaaggttgggcttcggaatcgtttccgggacggcgctggatgctcctcagcgcggggtatc  
catgctggagttcttcccccacccaacttgttattgcagcttataatggttacaataaagcaatagcatcacaatttcaca  
aataaagcattttttactgcattctagtgtggttgcacaaactcatcaatgtatcttatcatgtctgtataccgtcgaccttag

5 ctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgttatccgctcacaattccacacaacatacagagccgg  
 aagcataaagtgtaaagcctgggggtgcctaatagtgagctaactcacattaattgcgttgcgctcactgcccgtttccagt  
 cgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcgggttgcgtattggcgctcttcc  
 gcttctcgtcactgactcgtcgcgtcgttcgttcggctgcggcagcgggtatcagctcactcaaaggcggtaatacgg  
 10 ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgttaaaaag  
 gccgctgtgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga  
 aacccgacaggactataaagataccaggcgtttccccctggaagtcctcgtgcgctctcctgttccgacctgccgctta  
 ccggatacctgtccgcttttctcccttcgggaagcgtggcgttttcaatgctcacgctgtaggtatctcagtcggtgtagg  
 tcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccaccgctgcgcttatccggtaactatcgtctga  
 15 gtccaacccggtaagacacgacttatgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggc  
 ggtgctacagagttctgaagtgggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagcc  
 agttaccttcggaaaaagagttgtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttgttgaagca  
 gcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaac  
 tcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaatac  
 20 taaagtatatatgagtaaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctattcgttc  
 atccatagttgcctgactccccgtcgtgtagataactacgatacgggagggtaccatctggccccagtgctgcaatgata  
 ccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccggaaggggccgagcgcagaagtgtgt  
 cctgcaactttatccgctccatccagctctattaattgttgcgggaagctagagtaagtagtccgagtaatagtttgcgca  
 acgttgttgcattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaa  
 25 ggcgagttacatgatccccatgttgcacaaaaagcggtagctccttcggtcctccgatcgttgcagaagtaagtggcc  
 gcagtgttatcactcatggttatggcagcactgcataattctctactgcatgccatccgtaagatgcttttctgtactggtga  
 gtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgctcaatacgggataataccgcg  
 ccacatagcagaactttaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgag  
 atccagttcgtatgaacccactcgtgcacccaactgatcttcagcatctttactttaccagcgtttctgggtgagcaaaaaca  
 30 ggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgtgaatactatactcttcttttcaatattattg  
 aagcatttatcagggtattgtctcatgagcggatacatattgaatgtatttagaaaaataaacaataagggttccgcgcaca  
 ttccccgaaaagtgccacctgacgtc

30 2) INFORMATION FOR SEQ. ID. NO.66: HRV16 D35A primer

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

45 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....32

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:66:

5 gtgtcttattcatcagctttaatcatataccg

(2) INFORMATION FOR SEQ. ID. NO:67:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 base pairs

15 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

25 (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....34

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:67:

30 gtgaaccaggatgatgctgggtgggaattattatg

2) INFORMATION FOR SEQ. ID. NO.68: pcDNA3-3XUb-Bla HRV16 (C106A)

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7563 base pairs

40 (B) TYPE:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

45

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.68:



gacggatcgggagatctccgatccctatggctgactctcagtacaatctgctctgatgcccatagttaagccagtatctg  
 ctccctgcttgtgtgttgagggtcgtgagtagtgccgagcaaaatttaagctacaacaaggcaagccttgaccgacaatt  
 gcatgaagaatctgcttagggtaggcgttttgcgtgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
 ctagtattaatagtaataattacggggcattagttcatagccatataatggagttccgcgttacataacttacggtaaatggc  
 5 ccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
 ttccattgacgtcaatgggtggactatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacggcc  
 cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagacgttatgggactttcctacttggcagtag  
 atctacgtatttagtcatcgtattaccatgggtgatgcgggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
 ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 10 aactccgccccattgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
 gaaccactgcttactggcttactgaaattaatagactcactatagggagaccaagccttgatatagaattcctgcagccccg  
 ggggatctaccatggaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
 15 cacggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
 gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
 cacggatctaccatggaaatcttctgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattg  
 agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatcttctgctgggaaacagctg  
 20 gaagatggacgcaccctgtctgactacaacatccagaagagtcaccctgcacctggtactccgtctcagaggtgtgcac  
 cacggatctaccatgggaaacttctgttctgcgtattgtgaccagtgagcaattacacaaagtcaaaagtgtaacaaggatataca  
 caaagccaaacacaccaaagccttgggtcccaagaccaccagagctgttcaatactcacatacacataaccaccaactaca  
 aattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagtgcacatgtatgtg  
 catgttggtaatctaataacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaac  
 25 atataccgaacaagcacacaaggtgatggttatattccaacatgtaattgactgaagctacataattactgcaaacacaaaaa  
 caggtactaccaatattatgtcacacctcatgactggatgatagatacaagagagtgaatattatccaaacatatccagtaca  
 atttactaataggtgaaggaccatgtgaaccaggtgatgctgggtgggaaattattatgcaaacatggagtgataggtattatta  
 cagcaggtggtgagggccatgttgactcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacc  
 agaaacgctgggtgaaagttaaagatgctgaagatcagttgggtgcacagtggtgttacatgaactggatctcaacagcg  
 30 gtaagatccttgagagtttccgcccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatc  
 ccgtattgacgccgggcaagagcaactcggtcgccgatacactattctcagaatgacttgggtgagtactaccagtcaca  
 gaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtccataaccatgagtataactgcggccaac  
 ttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcg  
 ttgggaaccggagctgaatgaagccataccaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttg  
 35 cgcaaaactatttaactggcgaactacttactctagcttcccggaacaattaatagactggatggagggcgataaagttgcag  
 gaccacttctgcctcggcccttccggctggctggttattgctgataaatctggagccggtgagcgtgggtctcgcggtatc  
 attgcagcactggggccagatggtaagccctcccgatcgtagttatctacacgacggggagtcaggcaactatggatgaa  
 cgaaatagacagatcgtgagataggtgctcactgattaagcatttgtaatctagagggccctattctatagtgacaccaa  
 atgctagagctcgtgatcagcctcactgtgcttctagtgtccagccatctgttgttggccctccccgtgcttccctgac  
 40 cctggaaggtgccactcccactgtccttcttaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctg  
 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggct  
 ctatggcttctgaggggaaagaaccagctggggcttaggggggtatccccacgcgcctgtagcggcgcaatgaagcgc  
 ggcggtgtggtgtgttacgcgcagcgtgaccgctacacttgcagcggccttagcggcgctccttctgcttctcccttcc  
 ttctgccacgttcggcgcttccccgtcaagctctaaatcggggcatccctttaggggtccgatttagtgctttacggcacct  
 45 cgaccccaaaaaacttgattaggggtgatggttcacgtagtggccatcgcctgatagacggttttcggccttgacgttgg  
 agtccacgttcttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggctctattctttgattataagggg  
 tttggggatttccggcctattggttaaaaaatgagctgatttaacaaaaatgaacgcgaattaattctgtggaatgtgtgcagt



2) INFORMATION FOR SEQ. ID. NO.69: pcDNA3-3XUb-Bla HRV16 (D35A)

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 7563 base pairs

(B) TYPE:

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.69:

15 gacggatcgggagatctcccgatccctatggcgaactctcagtacaatctgctctgatgccgcatagtaagccagtatctg  
ctccctgcttggtgttgaggctcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
ctagtattaatagtaataatcattacggggcattagttcatagcccataatggagtccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
20 ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtacggcc  
cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagcatgaccttatgggactttcctacttggcagtag  
atctactgtattagtcacgctattaccatgggtgatcggttttggcagtagcatcaatgggcgtggatagcgggttgactcacgg  
ggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
25 gaacccactgcttactggcttatcgaataatacgaactactataggagacccaagcttgatcgaattcctgcagcccg  
ggggatctaccatgaaatcttcgtgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttcgtgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
30 agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
cacggatctaccatggaaatcttcgtgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattg  
agaatgtcaaggcaaaagatccaagacaagggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctg  
gaagatggacgcaccctgtctgactacaacatccagaaagagtcaccctgcacctggtagtccgtctcagaggtgtgcac  
35 cacggatccatgggaactttgttgcgtattgtgaccagtgagcaattacacaaagtcaaaagtggtaacaaggatataatca  
caaagccaaacacaccaaagcttggtgcccaagaccacccagagctgttaatactcacatacacataccaccaactaca  
aattgagttcagaagtacacaatgatgtggctataagacctagaacaatctaacaactgttgggcctagtacatgtatgtg  
catgttgtaataatatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagcttaatc  
atataccgaacaagcacacaaggatggttatattccaacatgtaattgactgaagctacatattactgcaaacacaaaaa  
40 caggtagtaccacaaatattgtcacacctcatgactggtagatgagatacaagagagtgaatattatccaaacatatccagtaca  
atttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgataggtattatta  
cagcaggtggtgagggccatgttcattcatagatcttagacactttcactgtgctgaaggatccggggcggtgctgcacc  
agaaacgctggtgaaagttaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcg  
gtaagatccttgagagtttgcggcgaagaacgtttccaatgatgagcacttttaagtctgctatgtggcgcggtattatc  
45 ccgtattgacggcgggcaagagcaactcggtcgccgcatcactattctcagaatgacttggtgagtagtaccagtcaca  
gaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtccataaccatgagtataactgcggccaac  
ttactctgacaacgatcgaggaccgaaggagtaaccgctttttgcacaacatgggggatcatgtaactgccttgatcg

ttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttg  
 cgaaactattaactggcgaactacttactctagcttccggcaacaattaatagactggatggaggcggataaagtgcag  
 gaccacttctgcgtcggcccttccggctggctggtttattgctgataaatctggagccggtagcgtgggtctcgcggtatc  
 attgcagcactggggccagatggttaagccctcccgtatcgtattctacacgacggggagtcaggcaactatggatgaa  
 5 cgaatatagacagatcgtgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaa  
 atgctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgttgtttgccccccccgtgccttcttgac  
 cctggaaggtgccactcccactgtccttcttaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctg  
 gggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggct  
 ctatggcttctgaggcggaaagaaccagctggggcttaggggggtatcccccacgcgccctgtagcggcgcaataagcgc  
 10 ggagggtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccttagcgcggccttctgcttcttcccttct  
 ttctgccacgttcgccggcttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgccttacggcacct  
 cgacccccaaaaacttgattagggtgatggttcacgtatggggccatcgccctgatagacggttttcgcctttgacgttgg  
 agtccacgttcttaatagtggactcttgtccaaactggaacaacactcaaccctatctcggctctattctttgattataaggga  
 ttttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatgaacgcgaattaattctgtggaatgtgtcagt  
 15 aggggtgtggaagtccccaggtccccaggcaggcagaagtatgcaaagcatgcactcaattagtcagcaaccaggtgt  
 ggaaagtccccaggtccccagcaggcagaagtatgcaaagcatgcactcaattagtcagcaaccatagtcggccct  
 aactccgccatccccgccctaactccgccagttccgccattctccgcccatggctgactaatttttttattatgcagag  
 gccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttcaaaaagctcc  
 cgggagcttgcctatccatttccggtatgatcaagagacaggatgaggatcgtttcgcattgaacaagatggattgcac  
 20 gcaggttctccggcgttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatccgc  
 cgtgttccggctgtcagcgcaggggcggcgggttcttttgcagaccacgttccgggtgccctgaatgaactgcagga  
 cgaggcagcgcggctatcgtggtggccacgacgggcgttcttgcgcagctgtgctcagctgtgactgaagcggga  
 agggactggctgctattgggcgaagtggcggggcaggatctcctgtcactcaccctgtcctgccgagaaagtatccatca  
 tggtgatcaatgcggcggctgcatacgttgatccggctacctgccattcgaccaccaagcgaacatcgcacgag  
 25 cgagcacgtactcggatggaagccggcttctgcagcaggtatgacgaagagcatcaggggctcgcgccagccg  
 aactgttcgccagggtcaaggcgcgcacgtcccgaggcaggtatcgtcgtgacccatggcgatgcctgcttgcgaat  
 atcatgttggaataaggcgttcttctgattatcgcactgtggcgggtgggtgtggcggaccgctatcaggacatagcgt  
 tggctacccgtgatattgtgaagagcttggcggcgaatgggtgaccgttctcgtgctttacggtatcggcgtcccgt  
 tcgagcgcacgccttctatgccttctgacgagttcttctgagcgggactctggggttgaaatgaccgaccaagcgac  
 30 gccaacctgccatcacgagatttgattccaccgccgccttctatgaaagggtgggttcggaatcgtttccgggacgcc  
 gggtggatgactcaccagcgcggggatctcatgctggagtcttcgccaccccaactgtttattgcagcttataatggttac  
 aaataaagcaatagcatcacaaatttcacaaataaagcatttttctactgcattctagtgtgtgttgcacaaactatcaatgta  
 tcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgttatccg  
 ctcacaattccacacaacatacagagccggaagcataaagtgtaaagcctgggggtgcctaatagtgagtaactacatta  
 35 attgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcgggg  
 agaggcgggttgcgtattgggcgcttctccgcttctcgtcactgactcgtcgcgtcggctgttggctgcggcgagcgg  
 tatcagctcactcaaggcggtaatacgttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc  
 cagcaaaaggccagggaaccgtaaaaaggccggttgcgttcttccataggctccgccccctgacgagcatcaca  
 aaatcgacgctcaagtgcaggtggcgaaaccgcagaggactataaagataaccaggcgttccccctggaagctccctg  
 40 tgcgtctcctgttccgacctgccgctaccggatacctgtccgccttctccttccgggaagcgtggcgcttctcaatgct  
 cacgctgtaggtatctcagttcgggtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagccgaccg  
 ctgcgccttatccggttaactatgctttagtccaacccggtgaagacacgacttatcgcactggcagcagccactggtaac  
 aggattagcagagcgaggtatgtaggcgggtgctacagagttctgaagtgggtggcctaactacggctacactagaaggac  
 agtatttggatctgcgctcgtgaagccagttaccttcggaaaaagagtggtagctcttgatccggcaacaaccaccg  
 45 ctggtagcgggtgtttttgttgcagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctac  
 ggggtctgacgctcagtggaacgaaaactcacgttaagggtatttgggtcatgagattatcaaaaggatcttcacctagatcc  
 ttttaataaaaaatgaagtttaataaatctaaagtatatatgagtaaaacttggtctgacagttaccaatgcttaacagtgagg



ctcatgactggtatgagatacaagagagtgaatattatccaaaacatatccagtacaattactaataaggtgaaggaccatgtg  
 aaccaggtgatgctggtgggaaattattatgcaaacatggagtgataggtattattacagcaggtggtgagggccatgttgc  
 attcatagatcttagacactttactgtgctgaaggatccggggcgtggctgacccagaaacgctggtgaaagtaaaagat  
 gctgaagatcagttgggtgcacgagtggttacatcgaactggatcacaacagcggtaagatccttgagagttttgccccg  
 5 aagaacgttttcaatgatgagcacttttaagttctgctatgtggcggtattatcccgtattgacggcgggcaagagcaa  
 ctggtgcggcgcatacactattctcagaatgacttggtgagtactaccagtcacagaaaagcatcttacggatggcatgac  
 agtaagagaattatgcagtgctgccataacctgatgtgataacactgaggccaacttacttctgacaacgatcggaggacc  
 gaaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagcc  
 ataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaactattaactggcgaactactt  
 10 actctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttcg  
 gctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtg  
 agccctcccgtatcgtatgtattctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagatag  
 gtgctcactgattaagcattggtaatctagaggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctc  
 actgtgcttctagtgtccagccatctgtgtttgccccccccgtgcttcttacccttggaggtgccactcccactgtcc  
 15 tttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagc  
 aaggggggaggttgggaagacaatagcaggcatgctggggatgcggtgggtctatggcttctgaggcggaaagaacc  
 agctggggctctagggggtatccccacgcgcctgtagcggcgcatgaagcgcggcggtgtggtgttacgcgcagcg  
 tgaccgtacacttgcagcgccttagcggcgctccttctgcttcttcccttcttctcgcacgttcggcggttccccg  
 tcaagcttcaaatcggggcatccctttaggggtccgatttagtgccttacggcacctcgaacccaaaaaacttgattagggtga  
 20 tggttcacgtagtggggcatcgcctgatagacgggttttgcgccttgacgttggagtcacgttcttaatagtgactctgt  
 tccaaactggaacaacactcaacctatctcggctattctttgattataagggttttggggttccggcctattggttaaaa  
 atgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgcagttagggtgtggaagtcgccaggtcccc  
 aggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaagtcgccaggtccccagcagg  
 cagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggcgcccttaactccgccatccccgcccttaactccg  
 25 cccagttccgccattctccgcccatggctgactaattttttattatgcagaggccgaggccgctctgcctctgagctat  
 tccagaagtagtgaggaggtttttggaggcctaggttttcaaaaagctccgggagcttgatatccatttccgatctg  
 atcaagagacaggatgaggatcgtttcgcgtatgtaacaagatggattgcacgcaggttctccggccgcttgggtggaga  
 ggctattcggctatgactgggcacaacagacaatcggctgctctgatccgccgtgtccggctgtcagcgcagggggcgc  
 ccggttcttttgaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggcc  
 30 acgacgggcgttcttgcgcagctgtgctcgcaggtgtcactgaagcgggaagggactggctgctattgggcgaagtgcc  
 ggggcaggatcctgtcctcacttgcctcctgccgagaaagtatccatcatggctgatgcaatgcggcggtgcatac  
 ctgtatccggctacctgccattcgaccaccaagcgaacatcgcacgagcagcagcgtactcggatggaagccggtctt  
 gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggtcaaggcgcgcacg  
 ccgacggcgaggatctcgtcgtgacctatggcgatgcctgcttgcgaatatcatggtggaatggccgctttctggatt  
 35 catcgaactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggtacccgtgatattgctgaagagcttgg  
 cggcgaaatgggctgaccgcttctcgtgctttacggatcgcgcgtcccgattcgcagcgcacgccttctatgccttctg  
 acgagtttcttgagcgggactctggggttcgaaatgaccgaccaagcgcgcccacctgccatcacgagatttcgattc  
 caccgccgcttctatgaaaggttgggcttcggaatcgtttccgggacggcggtggtgatcctccagcgcggggtatc  
 catgctggagttcttcgccaccccaactgtttattgcagcttataatggttacaataaagcaatagcatcacaatttcaca  
 40 aataaagcatttttactgcattctagtgtgtgttgcctaaactcatcaatgtatcttatcatgtctgtataccgtcgaccttag  
 ctgagcttggcgtaatatcatggtcatagctgttctctgtgaaattgttatccgctcacaaatccacacaacatacagccgg  
 aagcataaagtgtaaagcctgggggtgcctaatgagtgaactacacattaattgcgttgcgctcactgccgcttccagt  
 cgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggttgcgtattggcgctcttcc  
 gcttctcgtcactgactcgtcgcctcgtcgttcggctgcggcgagcgggtatcagctcactcaaggcggttaatacgg  
 45 ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag  
 gccgcgttgctggcggttttccataggtccgccccctgacgagcatcacaataatcagcgtcaagtcagaggtggcga  
 aaccggacaggactataaagataaccaggcgtttccccctggaagctccctgtgcgtctcctgttccgacctgccgcta



ccggatacctgtccgcctttctccctcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcgggtgtagg  
 tctgtcgtccaagctgggctgtgtgcacgaacccccgttcagcccaccgctgcgccttatccggttaactatcgtcttga  
 gtccaacccggtaagacacgacttatcgccactggcagcagccactggaacaggattagcagagcgaggtatgtaggc  
 ggtgtacagagttctgaagtgggtgcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagcc  
 5 agttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcgggtgtttttgttgcaagca  
 gcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggtctgacgctcagtggaacgaaaac  
 tcacgttaagggttttggcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaatcaatc  
 taaagtatatatgagtaacttggctgtgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttc  
 atccatagttgcctgactccccgtcgtgtgtagataactacgatacgggagggccttaccatctggccccagtgctgcaatgata  
 10 ccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccggaaggggcgagcgcagaagtgtg  
 cctgcaactttatccgctccatccagttctattaattgttgcgggaagctagagtaagtagttcggcagtaatagtttgcgca  
 acgttgttgcattgctacagcctcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccgggtcccaacgatcaa  
 ggcgagttacatgatccccatgttgcaaaaaagcggttagctccttcggctcctccgatcgttgcagaagtaagttggcc  
 gcagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtga  
 15 gtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggataataccgcg  
 ccacatagcagaactttaaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgag  
 atccagttcgtatgaacccactcgtgcacccaactgatcttcagcatctttactttaccagcgtttctgggtgagcaaaaaca  
 ggaaggcaaaatgccgcaaaaaagggaataaggggcgacacggaaatgttgaatactatactcttcttttcaatattattg  
 aagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaataagggttccgcgcaca  
 20 ttccccgaaaagtgccacctgacgtc

## 2) INFORMATION FOR SEQ. ID. NO.71: pcDNA3-Ub-Met-Bla HRV16 (D35A)

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7053 base pairs

(B) TYPE:

(D) TOPOLOGY: linear

### (ii) MOLECULE TYPE:

#### (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.71:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgcccatagttaagccagtatctg  
 ctccctgcttgtgtgttgagggtcgtgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaatt  
 40 gcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacggccagatatacgcgttgacattgattattga  
 ctagttattaatagtaaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc  
 ccgcttggtgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact  
 ttccattgacgtcaatgggtggactatttacggtaaactgccacttggcagtagatcaagtgtatcatatgccaagtacgcc  
 cctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagatgaccttatgggactttcctacttggcagtag  
 45 atctacgtattagtcacgtcattaccatgggtgatcggttttggcagtagatcaatgggcgtggatagcgggttgactcacgg  
 ggatttccaagtctccacccattgacgtcaatgggagtttgggttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
 aactccgccccattgacgcaaatgggcggttaggcgtgtacgggtggaggtctatataagcagagctctctggctaactaga

gaaccactgcttactggttactcgaattaatacgaactactataggagagaccaagcttggtaccaccatggagatcttcg  
 tgaagactctgactggaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaagatccaa  
 gacaaggaaggcatccctctgaccagcagaggtgatcttctgctgggaaacagctggaagatggacgcaccctgtctga  
 ctacaacatccagaaagagtcaccctgcacctggtactccgtctcagaggtgggatgcacggatccatgggaactttgtgt  
 5 tcgctgattgtgaccagtgaagcaattacacaaagtcaaagtggtaacaaggatatacacaaagccaaacacaccaaagctt  
 ggtgcccgaagaccaccagagctgttcaatactcacatacacataccaccaactacaattgagttcagaagtacacaatga  
 tgtggctataagacctaagaacaaactaacaactgttggcctagtgcacatgtatgtgcatgttggtatctaataacagaaat  
 ctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagctttaatcatataccgaacaagcacacaaggtg  
 atggttatattccaacatgtaattgcactgaagctacatactactgcaaacacaaaaacaggtactaccaattaatgtcacac  
 10 ctcatgactggtatgagatacaagagagtgaatattatccaaaacataccagtaacattactaataaggtgaaggaccatgtg  
 aaccaggtgattgtggtgggaaattattatgcaaacatggagtgaaggtattattacagcaggtggtgagggccatgttgca  
 ttcatagatcttagacacatttactgtgctgaaggatccggggcgtggctgacccagaacgctggtgaaagtaaaagatg  
 ctgaagatcagttgggtgcacgagtggtgtacatcgaactggatctcaacagcggtaagatccttgagagtttgcggccga  
 agaagctttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacggcgggaagagcaact  
 15 cggctgccgcatacactattctcagaatgacttgggtgagtactaccagtcacagaaaagcatcttaccggatggcatgaca  
 gtaagagaattatgcagtgtccataaccatgagtgaacactgcggccaacttacttctgacaacgatcggaggaccg  
 aaggagctaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagcca  
 taccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacggttgcgcaactattaactggcgaactactta  
 ctctagcttcccggcaacaattaatagactggatggaggcgggataaagttgcaggaccacttctgcgctcggcccttccgg  
 20 ctggtggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaa  
 gccctcccgtatcgtattatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagatagg  
 tgcctcactgattaagcattgtaacttagagggccctattctatagtgtcacctaaatgctagagctcgtgatcagcctcga  
 ctgtgccttctagtgtccagccatctgttgttggccctccccgtgccttccttgacctggaaggtgccactccactgtcct  
 ttctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagca  
 25 aggggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatggcttctgagggcggaaagaacca  
 gctggggctctagggggatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgt  
 gaccgtacacttggcagcggccatgcgcgccgtccttctgcttcttcccttcttctcggcacgtcggcggttcccccgt  
 caagctctaatacggggcatcccttttaggggtccgatttagtgccttacggcacctcgaccccaaaaaacttgattagggatg  
 ggttcacgtagtggggcatcggccctgatagacggttttgcgccttgacgttggagtcacgttcttaataaggacttctgtt  
 30 ccaactggaacaacactcaaccctatctcggctattctttgattataagggttttggggatttcggcctattggttaaaaa  
 atgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgcagttagggtgtggaaagccccaggctcccc  
 aggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcggcccccctaacccgcccaccccggccctaacccg  
 ccaagtgatgcaaagcatgcatctcaattagtcagcaaccatagtcggcccccctaacccgcccaccccggccctaacccg  
 35 tccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaaagctccgggagcttgatatccatttccgatctg  
 atcaagagacaggatgaggatcgtttcgcattgaacaagatggattgcacgcaggttctccggccgcttgggtggaga  
 ggctattcggctatgactgggcacaacagacaatcggctgctctgatccgccgtgttccggctgtcagcgcagggggcgc  
 ccggttcttttgcagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggcc  
 acgacgggcgttcttgcgcagctgtgtcgcaggtgtcactgaagcgggaagggactggctgctattggcggaagtgcc  
 40 ggggcaggatctcctgtcatctcacctgtcctgccgagaaagtatccatcatggtgatgcaatgcggcgggtgcatacg  
 ctgatccggctacctgcccattcgaccaccaagcgaacatcgcacgcagcagcactcggatggaagccgggtctt  
 gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttccagggtcaaggcgcgcagtc  
 ccgacggcgaggatctcgtcgtgacctggcgatccctgcttggcgaatatcatggtggaaaatggccgcttttctggatt  
 catcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttgg  
 45 cggcgaatgggctgaccgcttctcgtgctttacggatcgcgcgtcccgtatcgcagcgcacgccttctatgccttcttg  
 acgagttcttctgagcgggactctgggggtcgaatgaccgaccaagcgcgcccacctgccatcacgagatttcgattc  
 caccgccgcttctatgaaaggttgggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcggggatct



catgctggagttcttcgccaccccaacttgttattgcagcttataatggttacaataaagcaatagcatcacaaatttcaca  
 aataaagcattttttcactgcattctagttgtggttgccaaactcatcaatgtatcttatcatgtctgtataccgtcgaccttag  
 cttagagcttggcgtaatcatggtcatagctgttctgtgtgaaattgttatccgctcacaaattccacacaacatacgagccgg  
 aagcataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgcgctcactgcccgctttccagt  
 5 cgggaaacctgtcgtgccagctgcattaatgaatggccaacgcgcggggagagggcgtttgcgtattgggcgctctcc  
 gcttctcgtcactgactcgtgcgctcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaaacgg  
 ttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag  
 gcccggtgctggtgttttccataggctcggccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga  
 aacccgacaggactataaagataaccaggcgtttcccccgtgaagtcctcgtgcgctcctgttccgacctgccgctta  
 10 ccggatacctgtccgctttctcccttcgggaagcgtggcgcttttcaatgctcacgctgtaggtatctcagttcggtgtagg  
 tctgtcgtccaagctgggctgtgtgcacgaacccccgttcagccccaccgctgcgcttatccggtaactatcgtctga  
 gtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggc  
 ggtgctacagagttctgaagtggtagcctaactacggctacactagaaggacagtatttggtatctgcgctcgtgaagcc  
 agttaccttcgaaaaagagttgtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttgttgaagca  
 15 gcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggtctgacgctcagtgaacgaaac  
 tcacgttaagggaatttggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatc  
 taaagtatatatgagtaacttggtctgacagttaccaatgcttaatcagttaggcacctatctcagcgtatcttattcgttc  
 atccatagttgctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtctgcaatgata  
 ccgcgagaccacgctcaccggctccagattatcagcaataaaccagccagccggaaggccgagcgcagaagtgggt  
 20 cctgcaactttatccgctccatccagtctattaattgttgcgggaagctagagtaagtagttcccggttaaatagtttgcgca  
 acgttgttgccattgctacaggcatcgtggtgtcacgctcgtcttggtagtggttcattcagctccggttcccaacgatcaa  
 ggcgagttacatgatccccatgttgtgcaaaaaagcggtagctccttcggtcctccgatcgttgcagaagtaagttggcc  
 gcagtgttatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgactggtga  
 gtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccggcgtcaatacgggataataccgcg  
 25 ccacatagcagaactttaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgag  
 atccagttcgtatgtaaccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaaca  
 ggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatgttgaatactcatacttctcttttcaatattattg  
 aagcatttatcagggttattgtctcatgagcggatacatattgaatgtatttagaaaaataaacaataagggttccgcgcaca  
 tttccccgaaaagtgccacctgacgtc  
 30

## 2) INFORMATION FOR SEQ. ID. NO.72: pcDNA3-MetUb-Bla HR14

### 35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7512 base pairs

(B) TYPE: nucleic acid

40

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

### 45 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7512

5

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:72:

gacggatcgggagatctccgatccctatggctgactctcagtacaatctgctctgatgccgcatagttaagccagtatctg  
ctccctgcttgtgtgttgaggctcgtgagtagtgcgcgagcaaaatctaagctacaacaaggcaaggcttgaccgacaatt  
gcatgaagaatctgcttagggtaggcgtttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattga  
10 ctagtattaatagtaataattacggggctcattatgcatagcccataatggagttccgcgttacataacttacggtaaatggc  
ccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgtcccatagtaacgccaatagggact  
ttccattgacgtcaatgggtggactattacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccc  
cctattgacgtcaatgacggtaaatggcccgcctggcattatgccagtagcatgaccttatgggactttctacttggcagtag  
atctacgtatttagtcacgtattaccatgggtgatgcggttttggcagtagcatcaatgggcgtggatagcggttgactcacgg  
15 ggatttccaagctctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac  
aactccgccccattgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactaga  
gaacccactgcttactgcttaccgaaattaatacgaactactataggagaccaagcttggtagccaccatggagatcttcg  
tgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaatgatccaa  
gacaaggaaggcatccctcctgaccagcagaggtgatcttctgctgggaaacagctggaagatggacgcacccctgtctga  
20 ctacaacatccagaagagtcacccctgcacctggtactccgtctcagaggtgggatgcacggatccttgggtcgtgcagc  
ttgtgtgcatgtaactgaaatacaaaaagatgctactggaatagataatcacagagaagcaaaattgttcaatgattgga  
aaatcaacctgtccagccttgcacacttagaaagaaactggaactcttacttattgttaggtttgattctgagtataccatactg  
ggcactgcatctcaacctgattcagcaaaactattcaagcaatttgggtggccaagccatgtatgtccacatggtgccccgaa  
atccaaaagagtgggcgattacacatggcaaatgcttcaaacccagtgatttctcaaggtgggggatacatcaaggttt  
25 agtgtgccttatgtaggattggcatcagcatataattgttttatgatggttactcacatgatgatgcagaaactcagtatggcat  
aactgttctaaccatattgggtagtagtgccattcagaatagtaaatgaacatgatgaacacaaaactcttgaagatcagagt  
ttatcacaggggcaagctcgttgaagcatggattccaagagcaccagagcactaccctacacatcaatagggcgcaaa  
attatcctaagaatacagaaccagtaattaagaagaggaaaggtgacattaaatcctatggtttaggacctaggtacgggtgg  
gatttatacatcaaatgttaaaataatgaattaccacttgatgacaccagaagaccaccataatctgatagcacctatccaaat  
30 agagatttagcaaatagtctcaacaggagacatgggtgcagaaacaataccacactgtaaccgtacatcaggtgttactattc  
cacatattacagaaagtattacccataatttgcgaaaagcccaccaacatctggattgaagggaagccctattacccaagta  
gatttcaagcaggagtgtgaaaggggttggccggcagagctaggagactgcggtgggattttgagatgcatacatggt  
cccattggattgttaacagctgaaggtatggatattgtttgttctgacatacgcagttggagtgtatcgagagggaacag  
ggatccggggcggtgctgcaccagaaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggtt  
35 acatcgaactggatcacaacgcggtaagatccttgagagtttgcggccgaagaacgtttccaatgatgagcacttttaa  
gttctgctatgtggcgcggtattatccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatga  
cttgggtgagtactaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgtctgcataacc  
atgagtataactgcggccaacttctgacaacgatcggaggaccgaaggagtaaccgctttttgcacaacatgg  
gggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatg  
40 cctgtagcaatggcaacaacgttgcgcaaaactattaactggcgaactacttacttagcttcccggcaacaattaatagactg  
gatggaggcgataaagtgcaggaccacttctgcgctcgcccttccggctgggtggttattgctgataaatctggagcc  
gggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggaagccctccgtatcgtagtattctacacgacg  
gggagtcaggcaactatgatgaacgaaatagacagatcgtgagataggtgcctcactgattaagcattggtaatctaga  
gggccctattctatagtgcacctaaatgctagagctcgtgatcagcctcactgtgccttctagtgtccagccatctgtgttt  
45 gccctccccgtgcttcttaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcacgcatt  
gtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcag  
gcatgtcggggatcggtgggtctatggcttctgaggcggaagaaccagctggggctctaggggtatccccacgcg

ccctgtagcggcgcattaagcgcggcggtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccttagcgc  
 ccgctccttfcgctttctcccttcttctgccacgttcgcggctttcccgtcaagctctaaatcggggcatcccttaggg  
 ttcgatttagtgctttacggcacctcgaccccaaaaacttgattaggggtgatggttcacgtatgggccatcgccctgata  
 gacgggttttcgcccttgacgttgaggtccacgttcttaatagtggactctgttccaaactggaacaacactcaacccatct  
 5 cggctctattctttgattataagggattttggggatttcggcctatttggttaaaaaatgagctgatttaacaaaaattaacgcga  
 attaatctgtggaatgtgtgtcagttagggtgtggaagtccccaggctccccaggcaggcagaagtatgcaaagcatgc  
 atcicaattagtcagcaaccaggtgtggaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcattcaatt  
 agtcagcaaccatagtcggccctaaactccgcccattccgcccctaaactccgcccagttccgcccattctccgcccattg  
 gctgactaatTTTTTatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtatgaggaggctttttgg  
 10 aggcctaggccttttgcaaaaagctccgggagcttgatatccatttcggatctgatcaagagacaggatgaggatcggttc  
 gcatgattgaacaagatggattgcacgcagggttctccggcgttggtggagaggctattcggctatgactgggcacaac  
 agacaatcggctgctctgatccgcccgtgtccggctgtcagcgcaggggcgcccgggtctttttgtcaagaccgacctgtc  
 cgggtgcccgtaatgaactgcaggacgaggcagcgcggctatcggtggtggccacgacgggcgttccttgccagctgtg  
 ctgcaggtgtcactgaagcgggaagggtggtgctattgggcgaagtgcggggcaggatctctgtcatctcacctt  
 15 gctcctgccgagaagtatccatcatggctgatgcaatcgggcggtgtcatalcgttgatccggctacctgccattcgacc  
 accaagcgaaacatcgatcgagcgcagcagcgtactcggatggaagccggtctgtcgtatcaggatgatctggacgaaga  
 gcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcgtatgccgagggcaggatctcgtcgtgac  
 ccatggcgatgcctgcttgcgaatatcatggtggaatggcgtttcttgattcatcactgtggccggctgggtgtg  
 cggaccgctatcaggacatagcgttggtaccctgtatattgtgaagagcttgccggcgatggggtgaccgcttctcgc  
 20 tgccttacgggtatcgccgctcccgttcgcagcgcacgccttctatcgcttcttgacgagttctctgagcgggactctggg  
 gtgcgaatgaccgaccaagcgacgccaacctgccatcacgagatttcgattccaccgcccttctatgaaagggtggg  
 ctteggaaactgtttccgggacggcggtggtgatcctccagcgcggggtatctatgctggagtcttcgccaccccaac  
 ttgtttattgcagcttataatggttacaataaagcaatagcatcacaaattcacaaataaagcattttttactgcattctagt  
 gtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttgccgtaatcatggtcatag  
 25 ctgtttcctgtgtgaaattgttatccgctcacaaattccacaaacatacagaccggaagcataaagtgtaaagcctggggtgc  
 ctaatgagtgagctaactcacattaattgcgttgcgtcactgcccgtttccagtcgggaaacctgtcgtgccagctgcat  
 atgaatcgcccaacgcgcggggagaggcggttgcgtattggcgcttctccgcttctcgtcactgactcgtcgcgtc  
 ggtcgttcggctgcggcgagcgggtatcagctcactcaaggcggttaatacgggtatccacagaatcaggggataacgcag  
 gaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaggccgcgttgcgtggcgttttccataggct  
 30 ccgccccctgacgagcatcacaaaaatcgacgctcaagtcagagtggtggcgaacccgacaggactataaagataaccag  
 gcgtttccccctggagctccctcgtgcgtctctgttccgacctgccgttaccggataacctgtccgcttttcccttcg  
 ggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcgggtaggtcgtcgtccaaagctgggctgtgtgca  
 cgaacccccgttcagcccagccgtcgccttatccggttaactatcgtcttgagtccaacccggttaagacacgacttatcg  
 ccactggcagcagccactggtaacaggattagcagagcaggtatgtaggcgggtgctacagagttcttgaagtgtggcc  
 35 taactacgggtacactagaaggacagtatttggtatctgcgtctgctgaagccagttaccttcggaaaaagagttggtagct  
 cttagctcggaacaaccaccgctggtagcgggtgtttttgttgcaagcagcagattacgcgcagaaaaaaggatc  
 tcaagaagatcctttgatctttctacggggtcgtgacgctcagtggaacgaaaactcacgttaagggtttttgtcatgagatta  
 tcaaaaaggatcttcacttagatccttttaaaataaaatgaagttttaaataatcaatcaaaagtatatatgagtaacttggtcgtac  
 agttaccaatgcttaatcagtgaggcacctatctcagcgaatctgtctatttcgttcacatagttgcctgactccccgtcgtga  
 40 gataactacgatacgggaggggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccag  
 atttatcagcaataaaccagccagccggaaggggcgcagcgcagaagtgtcctgcaactttatccgctccatccagctat  
 taattgttggcgggaagctagagtaagtagttcggcagttaatgatttgcgcacgttgttgcattgctacaggcatcgtggt  
 gtcacgctcgtcgttggatggcttcaatcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaa  
 aaaagcggtagctccttcgtcctccgacgttgcagaaagtaagttggccgcagttatcactcatggttatggcagcact  
 45 gcataattctcttactgtcatgccatccgtaagatgctttctgtgactggtgagttactcaaccaagtcattctgagaatagtgat  
 gcggcgaccgagttgctcttggccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatcatt  
 ggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaacccactcgtgcacca

actgatcttcagcatcttttactttaccagcgtttctgggtgagcaaaaacaggaaggcaaatgccgaaaaaaggggaat  
aagggcgacacggaaatgttgaatactcatactcttcttttcaatattattgaagcattatcagggttattgtctcatgagcg  
gatacatatttgaatgtatttagaaaaataaacaatataggggtccgcgcacatttccccgaaaagtgccacctgacgtc

5

10

15

actgatcttcagcatcttttactttaccagcgtttctgggtgagcaaaaacaggaaggcaaatgccgaaaaaaggggaat  
aagggcgacacggaaatgttgaatactcatactcttcttttcaatattattgaagcattatcagggttattgtctcatgagcg  
gatacatatttgaatgtatttagaaaaataaacaatataggggtccgcgcacatttccccgaaaagtgccacctgacgtc

## CLAIMS

We claim.

1. A method of detecting an activity in a cell, comprising;

5           1) providing a cell comprising,

a) at least one destabilization domain, wherein said destabilization domain is non-cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases,

b) a reporter moiety, and

10           c) a linker moiety that operatively couples said destabilization domain to said reporter moiety,

wherein said linker moiety comprises a recognition motif for said activity and modification of said linker moiety by said activity modulates the coupling of said destabilization domain to said reporter moiety thereby modulating the stability of said reporter moiety, and

15           wherein said linker moiety is non-cleavable by said  $\alpha$ -NH-ubiquitin protein endoproteases,

2) detecting said reporter moiety, or a product of said reporter moiety.

20

2. The method of claim 1, wherein said at least one destabilization domain is arranged as linear multimer, and

wherein said linear multimer comprises at least two copies of said destabilization domain and is non-cleavable by said  $\alpha$ -NH-ubiquitin protein endoproteases.

25

3. The method of claim 1, wherein said linker moiety is non-naturally occurring polypeptide or protein.

30   4. The method of claim 1, wherein said linker moiety covalently couples said destabilization domain to said reporter protein.

5. The method of claim 1, wherein said linker moiety is between about 1 and 30 amino acid residues.
6. The method of claim 1, wherein said destabilization domain comprises a ubiquitin homolog.
7. The method of claim 6, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by said  $\alpha$ -NH-ubiquitin protein endoproteases.
8. The method of claim 6, wherein said ubiquitin homolog comprises a mutation at glycine 76.
9. The method of claim 1, wherein said linker moiety comprises a first amino acid sequence that is covalently coupled to said reporter moiety, and a second amino acid sequence that is covalently coupled to said at least one destabilization domain.
10. The method of claim 1, wherein said activity is selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.
11. The method of claim 1, wherein said reporter moiety is selected from the group consisting of a naturally fluorescent protein homolog, a  $\beta$ -lactamase homolog, a  $\beta$ -galactosidase homolog, an alkaline phosphatase homolog, a CAT homolog, and a luciferase homolog.
12. The method of claim 11, wherein said reporter moiety comprises a  $\beta$ -lactamase homolog.
13. The method of claim 11, wherein said reporter moiety comprises an *Aequorea* Green fluorescent protein homolog.

14. The method of claim 11, wherein said reporter moiety comprises an Anthozoan Green fluorescent protein homolog.
15. The method of claim 1, wherein said cell is a mammalian cell.
- 5 16. The method of claim 1, wherein said cell is a yeast cell.
17. The method of claim 1, wherein said cell is an insect cell.
- 10 18. The method of claim 1, wherein said cell is a plant cell.
19. The method of claim 1, wherein said method further comprises the step of adding a protein synthesis inhibitor to said cell.
- 15 20. The method of claim 1, wherein said method further comprises the step of adding an inhibitor of said reporter moiety to said cell.
21. The method of claim 1, wherein said method further comprises the step of adding a test chemical to said cell.
- 20 22. The method of claim 20, wherein said method further comprises the step of relating said reporter moiety activity before addition of said test chemical to said reporter moiety activity after addition of said test chemical.
- 25 23. A method of regulating the concentration of one or more target proteins in a cell, comprising;
- 1) providing a cell comprising,
- a) a linear multimerized destabilization domain, wherein said linear multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises
- 30 at least two copies of a destabilization domain,
- b) a target protein, and

c) a linker that operatively couples said linear multimerized destabilization domain to said target protein,

wherein said linker comprises a protease cleavage site for a protease and cleavage of said linker by said protease modulates the coupling of said linear multimerized destabilization domain to said target protein, thereby modulating the stability of said target protein in said cell, and

wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases,

2) providing said protease to cause cleavage of said linker thereby increasing the stability and concentration of said protein of interest in said cell.

24. The method of claim 23, wherein said protease is naturally expressed in said cell.

25. The method of claim 23, wherein said protease is not naturally expressed in said cell.

26. The method of claim 23, further comprising the step of adding an inhibitor of said protease.

27. The method of claim 23, wherein said linker is between 1 and 30 amino acid residues.

28. The method of claim 23, wherein said cell is a mammalian cell.

29. The method of claim 23, wherein said cell is a yeast cell.

30. The method of claim 23, wherein said cell is an insect cell.

31. The method of claim 23, wherein said destabilization domain comprises a ubiquitin homolog.



32. The method of claim 31, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.
33. The method of claim 31, wherein said ubiquitin homolog comprises a mutation at  
5 glycine 76.
34. The method of claim 23, wherein said protease is provided by transfecting said cell with an expression vector comprising a nucleic acid sequence encoding said protease.
- 10 35. The method of claim 34, wherein said expression vector further comprises an inducible promoter.
36. The method of claim 34, wherein said expression vector is a retroviral expression  
15 vector.
37. The method of claim 34, wherein said protease is a viral protease.
38. A method of destabilizing a target protein in a cell, comprising;  
20 operatively coupling a target protein to a linear multimerized destabilization domain, wherein said linear multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of a destabilization domain.
- 25 39. The method of claim 38, wherein said destabilization domain comprises a ubiquitin homolog.
40. The method of claim 39, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.
- 30 41. The method of claim 39, wherein said ubiquitin homolog comprises a mutation at glycine 76.

42. The method of claim 38, wherein said protein of interest is fused in frame to said multimerized destabilization domain.
- 5 43. The method of claim 38, wherein said protein of interest is non-covalently coupled to said multimerized ubiquitin fusion protein.
44. The method of claim 38, wherein said cell is a mammalian cell.
- 10 45. The method of claim 38, wherein said cell is a yeast cell.
46. The method of claim 38, wherein said cell is an insect cell.
47. The method of claim 38, wherein said cell is a plant cell.
- 15 48. The method of claim 38, wherein said target protein is coupled to said multimerized destabilization domain by a linker.
49. The method of claim 48, wherein said linker is between 1 and 10 amino acid  
20 residues.
50. A recombinant DNA molecule, comprising a nucleic acid sequence encoding for;
- 25 a) a linear multimerized destabilization domain, wherein said linear multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of a destabilization domain,
- b) a target protein, and
- c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,
- 30 wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases.

51. The method of claim 50, wherein said linker moiety comprises an enzyme modification site for an activity, and modification of said linker moiety by said activity modulates the coupling of said multimerized destabilization domain to said reporter moiety, thereby modulating the stability of said reporter moiety.

5

52. The method of claim 50, wherein said destabilization domain comprises a ubiquitin homolog.

53. The method of claim 52, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.

10

54. The method of claim 52, wherein said ubiquitin homolog comprises a mutation at glycine 76.

55. A recombinant protein molecule, comprising an amino acid sequence encoding for;

15

a) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

20

b) a target protein, and

c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

25

wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases.

56. The method of claim 55, wherein said linker moiety comprises a recognition motif for an activity, and modification of said linker moiety by said activity modulates the coupling of said multimerized destabilization domain to said reporter moiety, thereby modulating the stability of said reporter moiety.

30

57. The method of claim 55, wherein said destabilization domain comprises a ubiquitin homolog.

58. The method of claim 57, wherein said ubiquitin homolog comprises a mutation  
5 that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.

59. The method of claim 57, wherein said ubiquitin homolog comprises a mutation at glycine 76.

10 60. A host cell, comprising a nucleic acid sequence encoding for;

a) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

15 b) a target protein, and

c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases.

20 61. The method of claim 60, wherein said destabilization domain comprises a ubiquitin homolog.

62. The method of claim 61, wherein said ubiquitin homolog comprises a mutation  
25 that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.

63. The method of claim 61, wherein said ubiquitin homolog comprises a mutation at glycine 76.

30 64. A transgenic animal, comprising a nucleic acid sequence encoding for;

d) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a

$\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

- e) a target protein, and
  - f) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,
- wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases.

5

65. The method of claim 64, wherein said destabilization domain comprises a ubiquitin homolog.

10

66. The method of claim 65, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.

67. The method of claim 65, wherein said ubiquitin homolog comprises a mutation at glycine 76.

15

68. A transgenic plant, comprising a nucleic acid sequence encoding for;

- a) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,
  - b) a target protein, and
  - c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,
- wherein said linker is non-cleavable by a  $\alpha$ -NH-ubiquitin protein endoproteases.

20

25

69. The method of claim 68, wherein said destabilization domain comprises a ubiquitin homolog.

30

70. The method of claim 69, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by  $\alpha$ -NH-ubiquitin protein endoproteases.

71. The method of claim 69, wherein said ubiquitin homolog comprises a mutation at  
5 glycine 76.

72. A method for identifying a modulator of an activity, comprising;

a) contacting a cell with a test chemical, wherein said cell comprises,

i) at least one destabilization domain, wherein said destabilization  
10 domain is non-cleavable by  $\alpha$ -NH-ubiquitin protein endoproteases,

ii) a reporter moiety, and

iii) a linker moiety that operatively couples said destabilization domain  
to said reporter moiety,

wherein said linker moiety comprises a recognition motif for  
15 said activity and modification of said linker moiety by said activity  
modulates the coupling of said destabilization domain to said reporter  
moiety thereby modulating the stability of said reporter moiety, and

wherein said linker moiety is non-cleavable by said  $\alpha$ -NH-  
ubiquitin protein endoproteases,

b) detecting said reporter moiety, or a product of said reporter moiety in the  
20 presence of said test chemical, and

c) comparing said reporter moiety activity from step b) to the reporter moiety  
activity in a control cell in the absence of said test chemical.

73. The method of claim 72, further comprising the step of contacting said cell with  
25 an activator of said activity prior to the addition said test chemical.

74. The method of claim 72, further comprising the step of detecting the viability of  
said cell  
30

75. The method of claim 72, wherein said activity is selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.
- 5 76. The method of claim 72, wherein said reporter moiety is selected from the group consisting of a naturally fluorescent protein homolog, a  $\beta$ -lactamase homolog, a  $\beta$ -galactosidase homolog, an alkaline phosphatase homolog, a CAT homolog, and a luciferase homolog.
- 10 77. A test chemical identified by any of the methods of claims 72, 73, 74, 75 or 76.
78. A pharmaceutical composition comprising a test chemical identified by any one of the methods of claims 72, 73, 74, 75 or 76.
- 15 79. The pharmaceutical composition of claim 78, further comprising a pharmaceutically acceptable carrier.

20

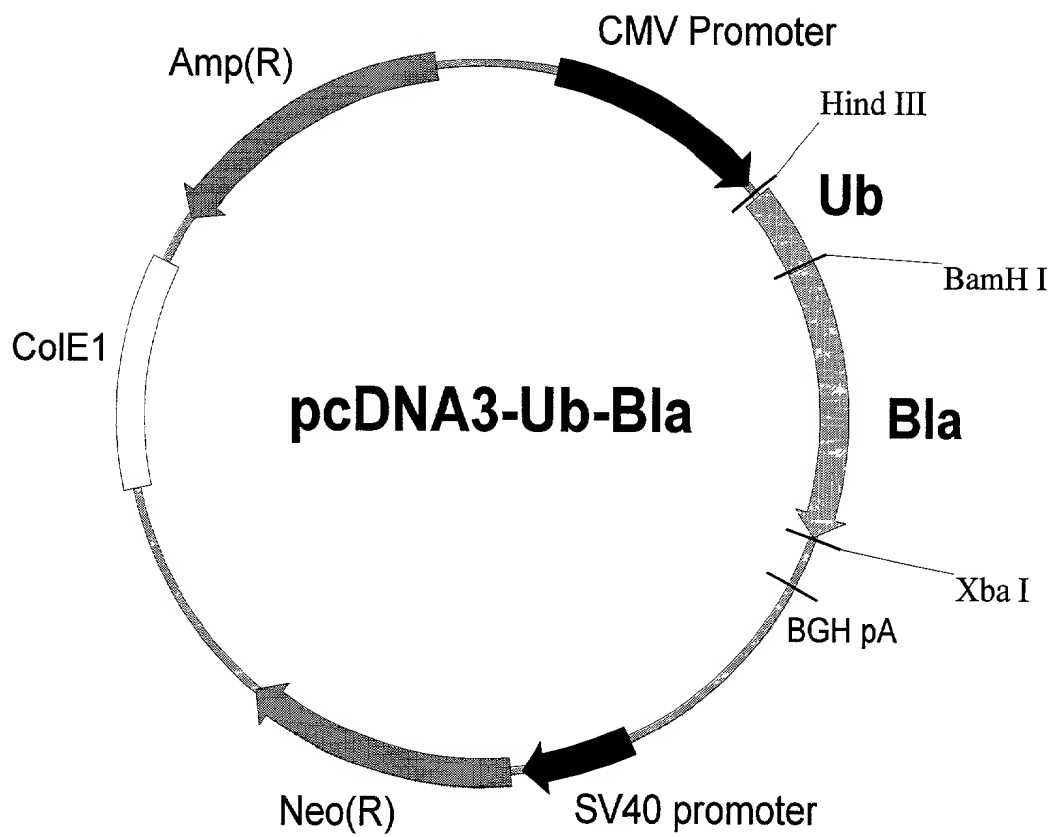
## ABSTRACT

### 5 METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF

This invention is directed towards methods of destabilizing proteins in living cells, and their use for the development of novel assays. In one embodiment, the invention  
10 comprises the use of non-cleavable multimerized ubiquitin fusion proteins to destabilize a target protein, such as a reporter moiety. In one aspect of this method the constructs also comprises a linker that operatively couples the reporter moiety to the multimerized ubiquitin fusion protein. In this embodiment, enzymatic modification of the linker results in a modulation of the coupling of the reporter protein to the  
15 multimerized ubiquitin domains resulting in a change in the stability of the reporter moiety. The level of the reporter moiety in the cell can then be used as a measure of the enzymatic activity in the cell. In another embodiment the invention provides for a generalized way of coordinately regulating the cellular concentration of a plurality of target proteins. In one aspect of this method, the target proteins are operatively  
20 coupled to a ubiquitin fusion protein via a linker containing a protease cleavage site. Cleavage of the linker by a protease results in uncoupling of the target protein from the multimerized ubiquitin construct, and results in an increase in the stability and concentration of the target protein.

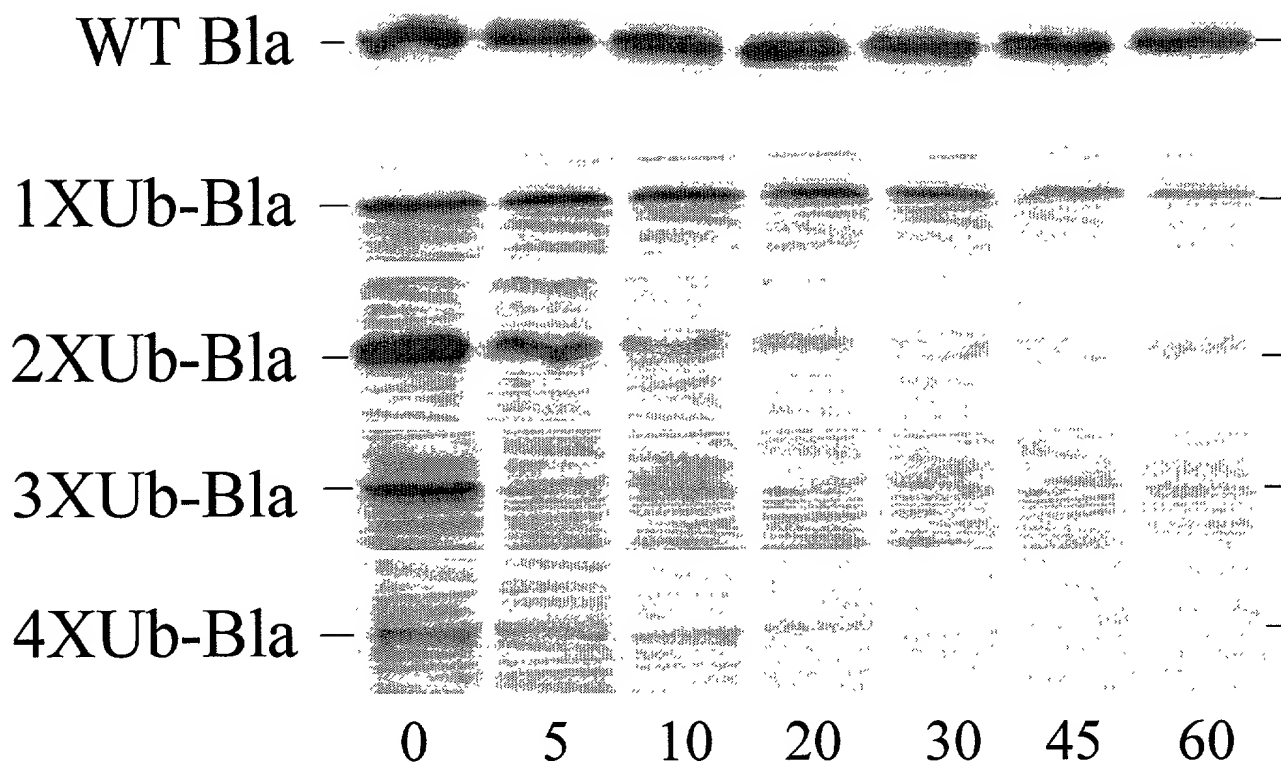


# Figure 1



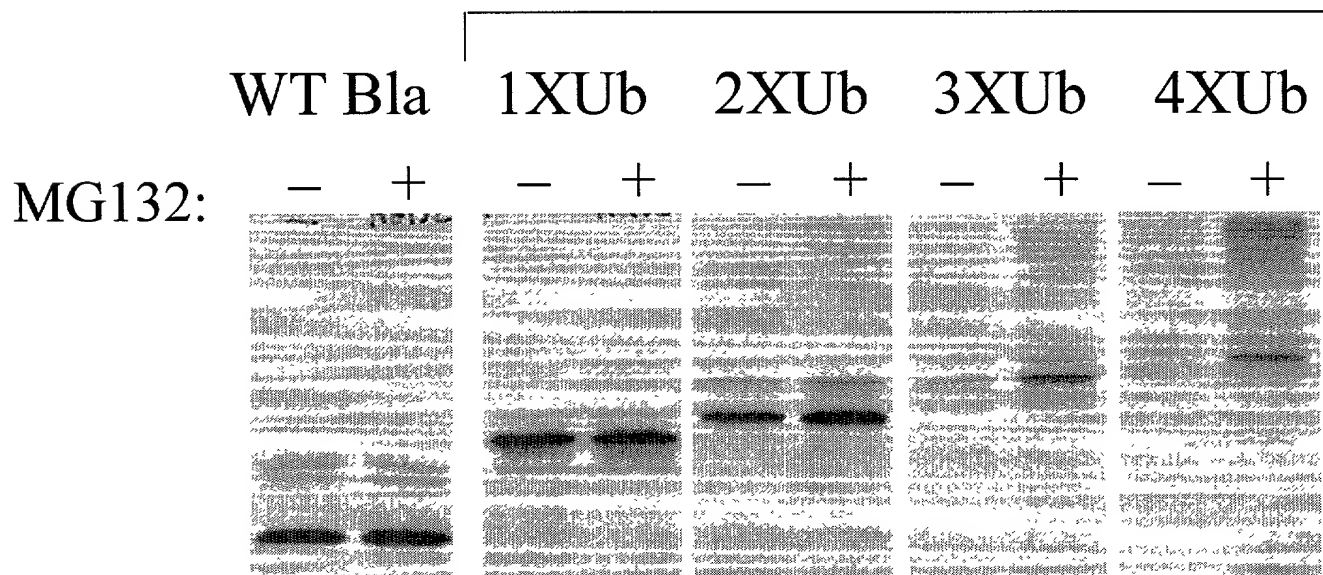
# Figure 2

**A.**

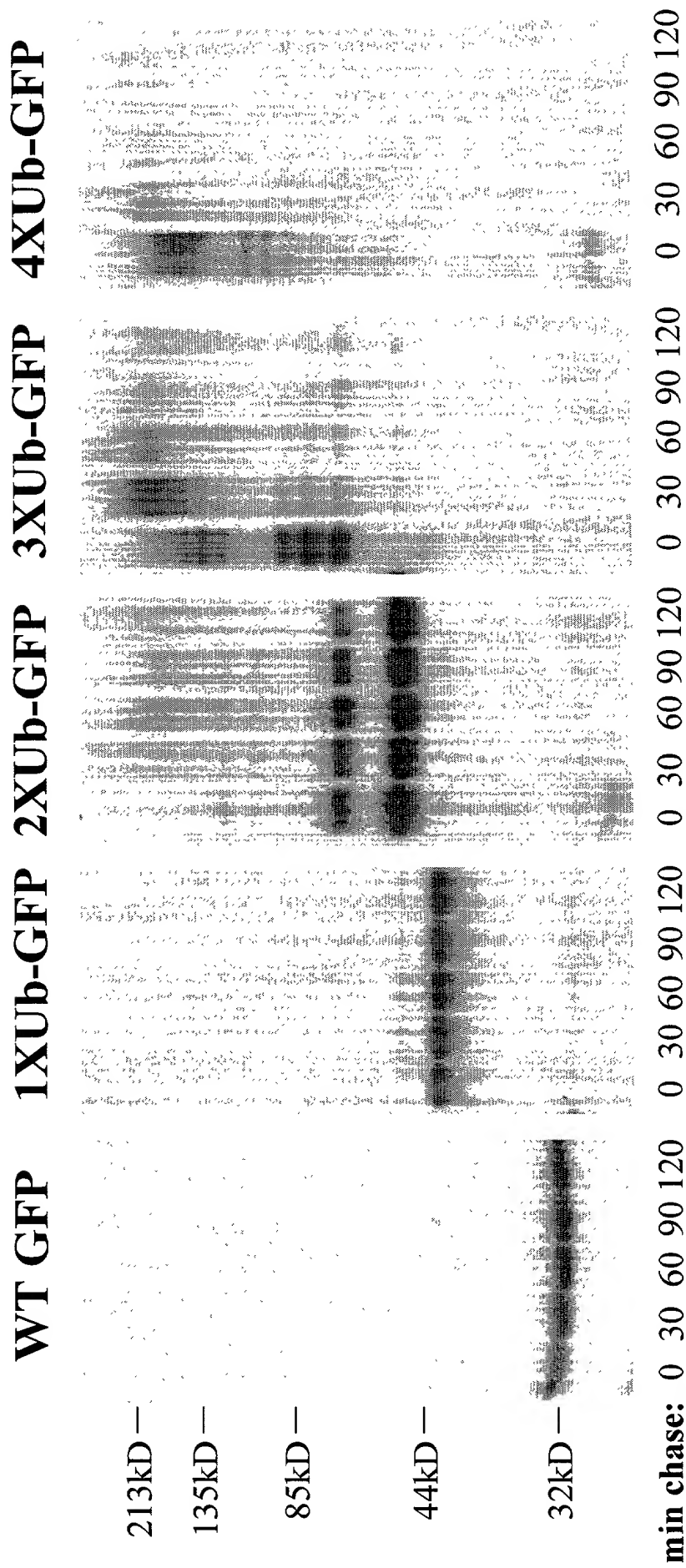


**B.**

UbG76V-Bla fusions



**Figure 3**



# Figure 4

WT Casp-3

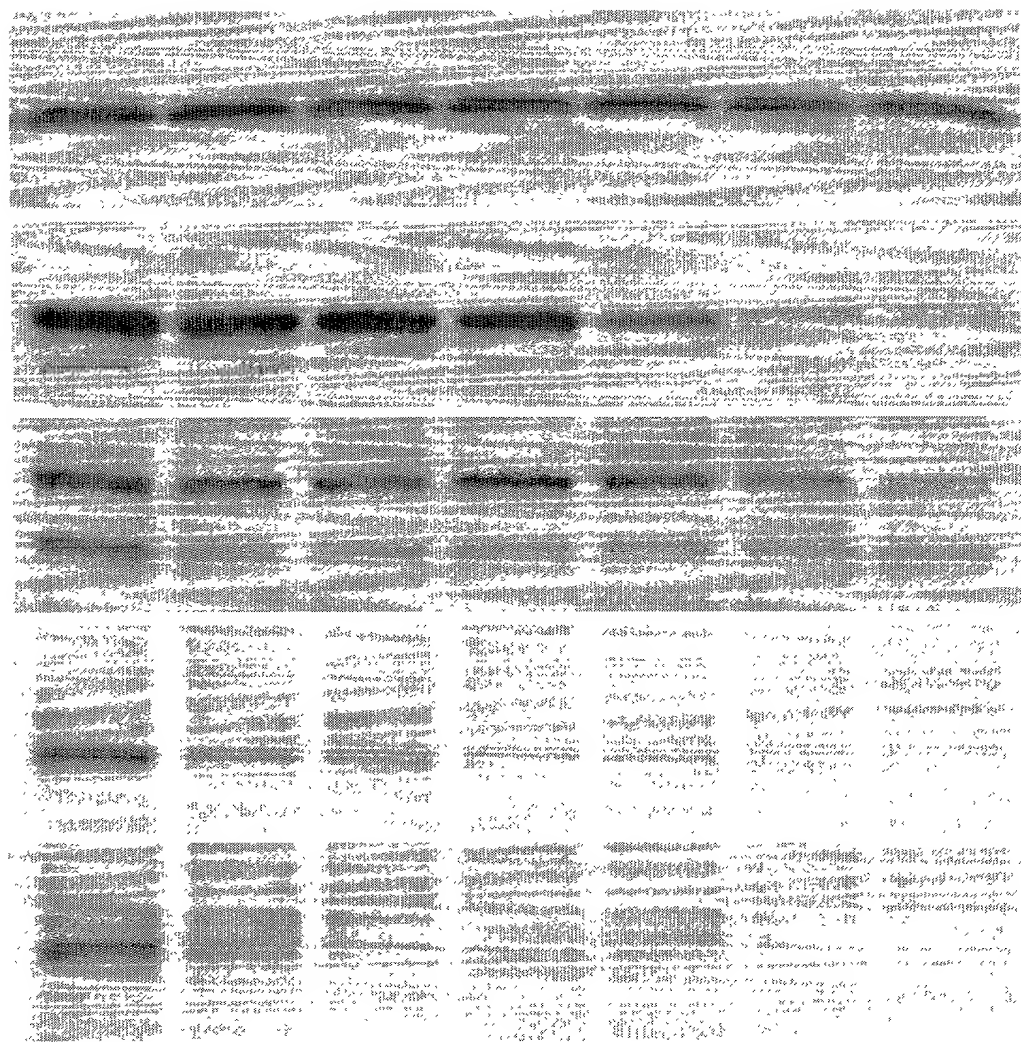
1XUb-C3

2XUb-C3

3XUb-C3

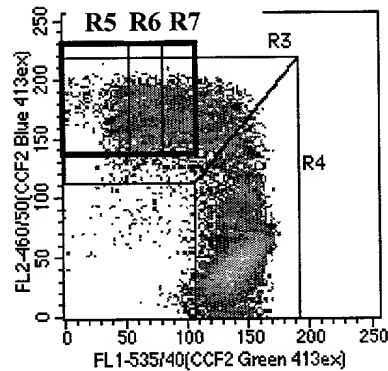
4XUb-C3

min chase: 0 5 10 20 30 45 60

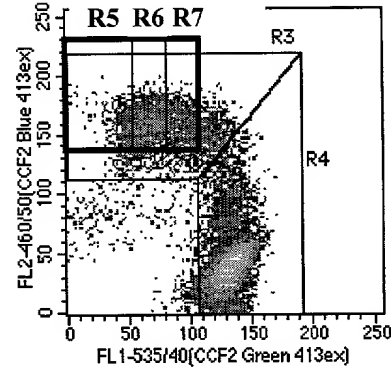


# Figure 5

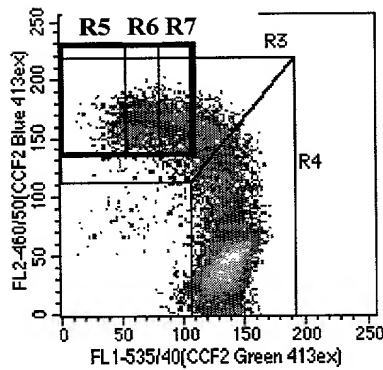
**WT Bla**



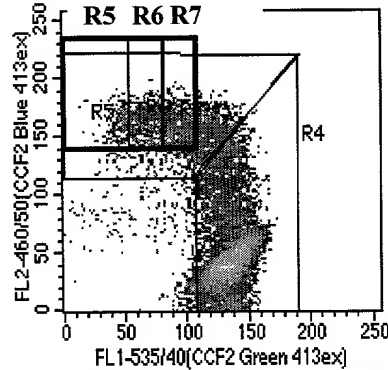
**1XUb-Bla**



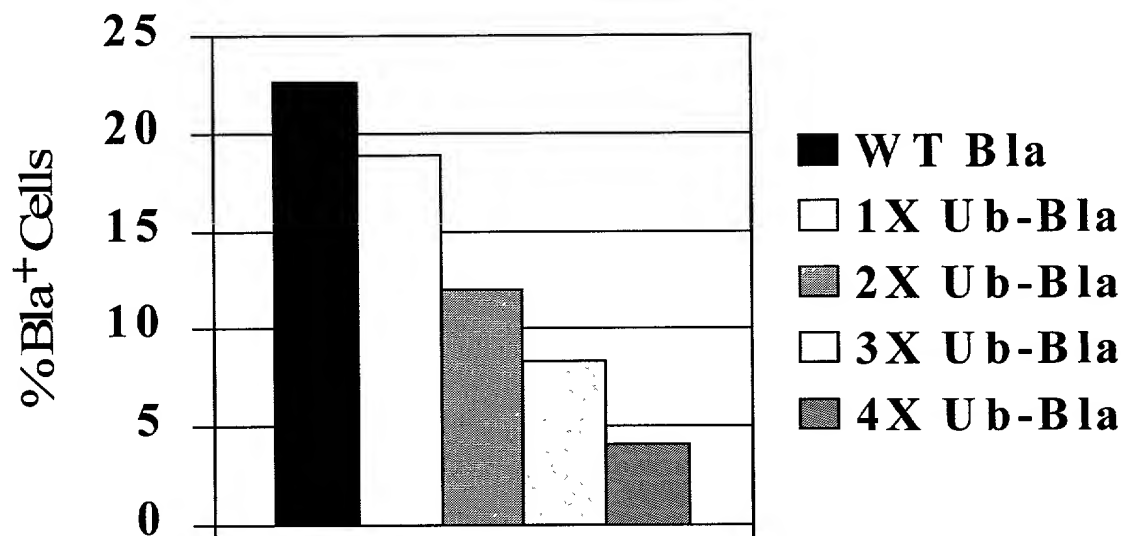
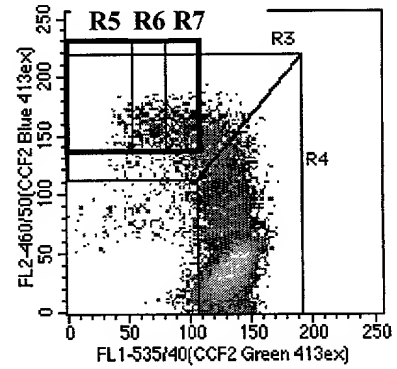
**2XUb-Bla**



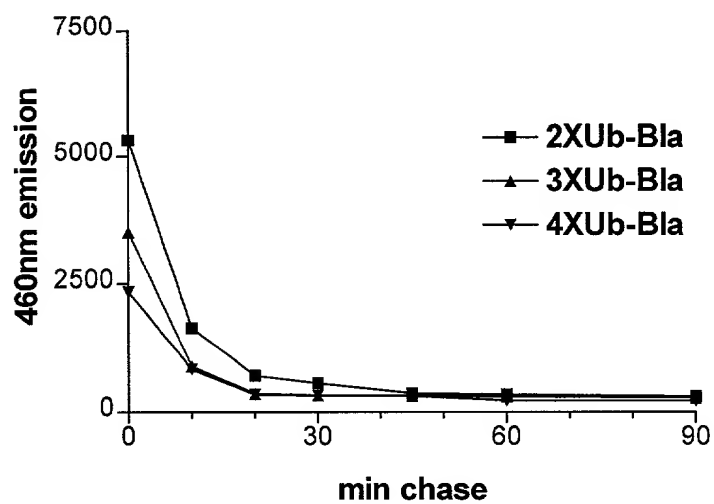
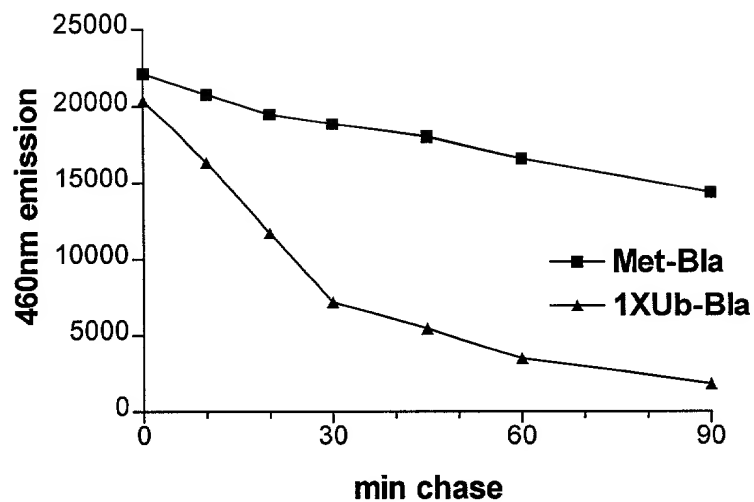
**3XUb-Bla**



**4XUb-Bla**

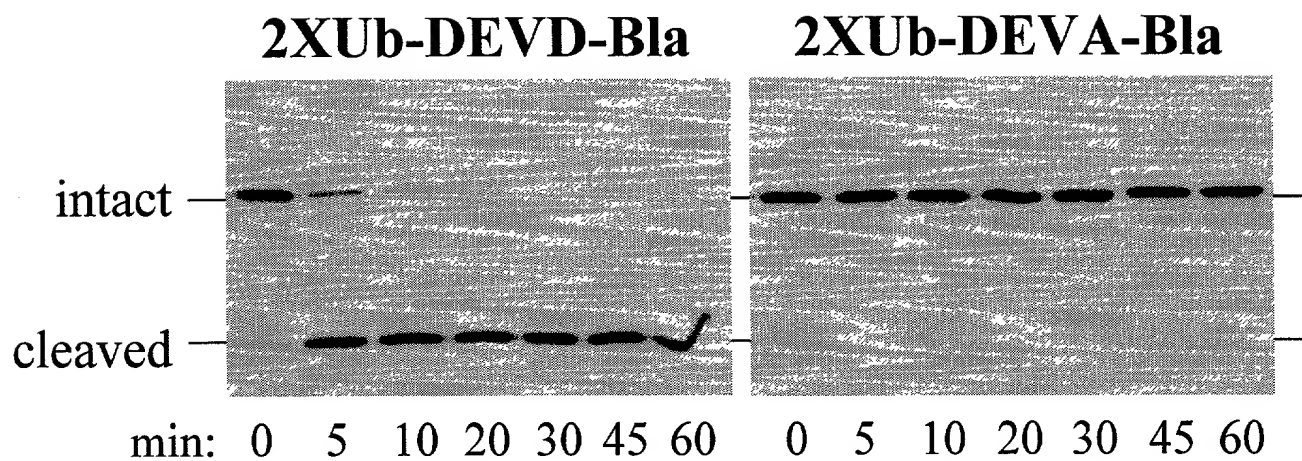


# Figure 6

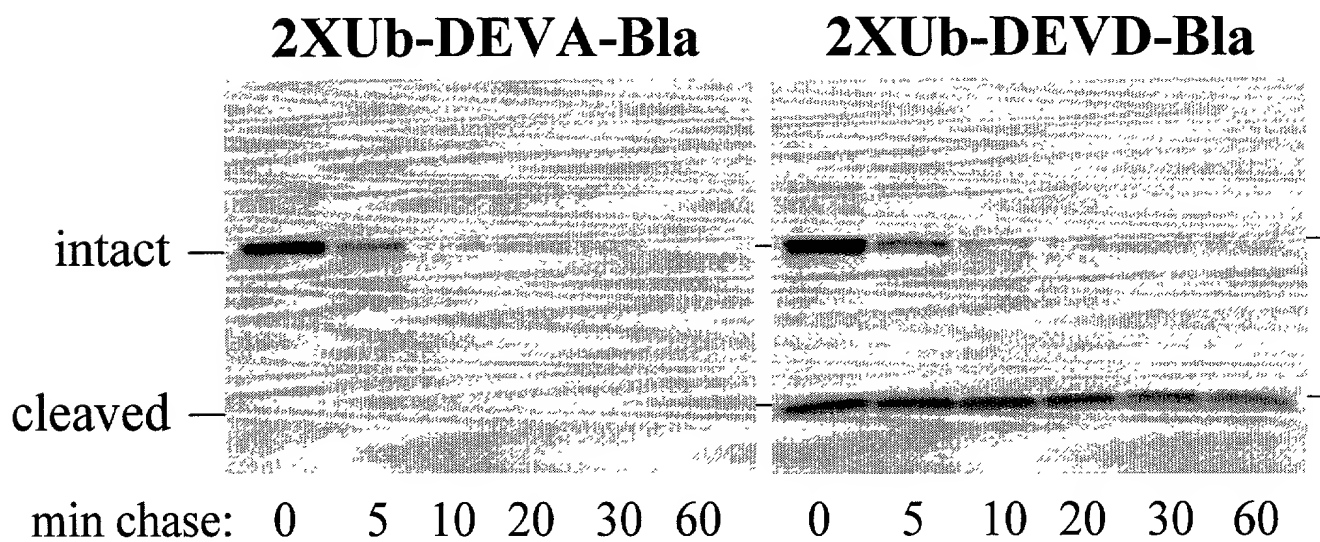


# Figure 7

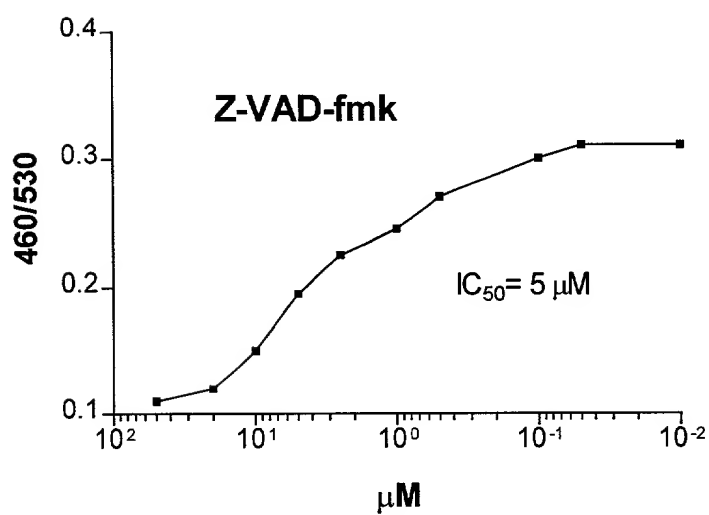
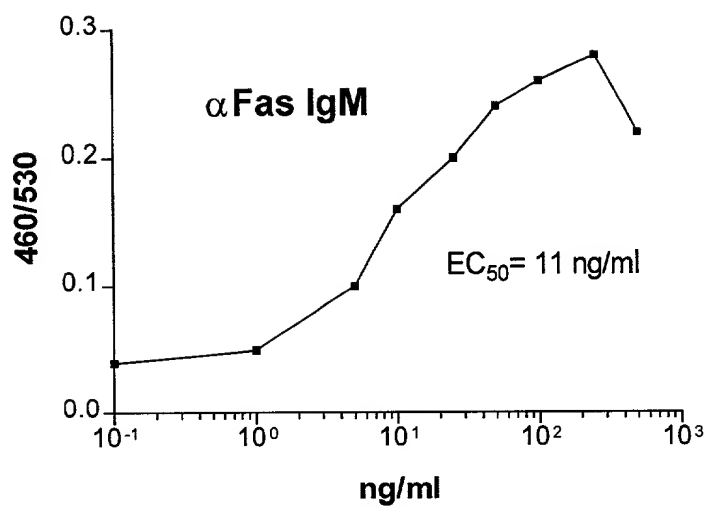
A.



B.



# Figure 8

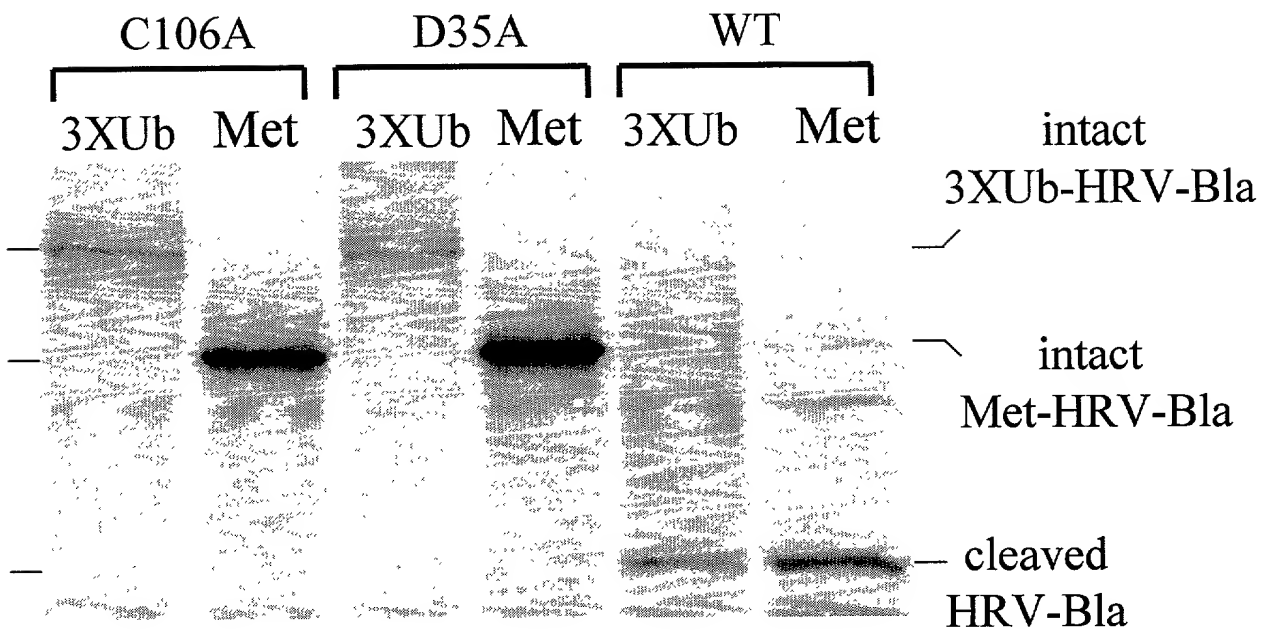




# Figure 9

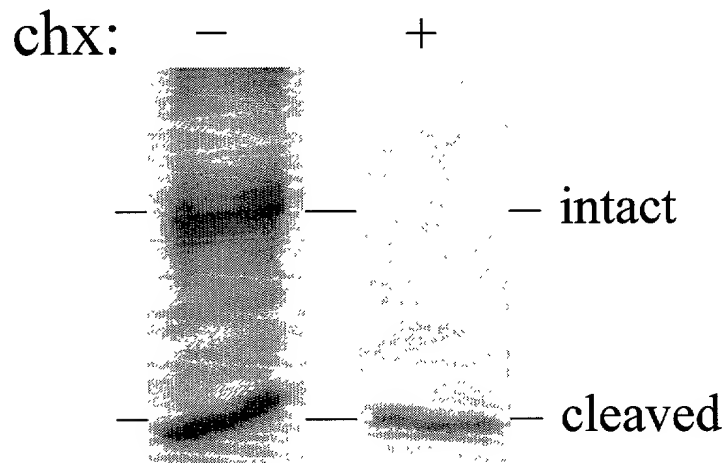
**A.**

## HRV16-Bla

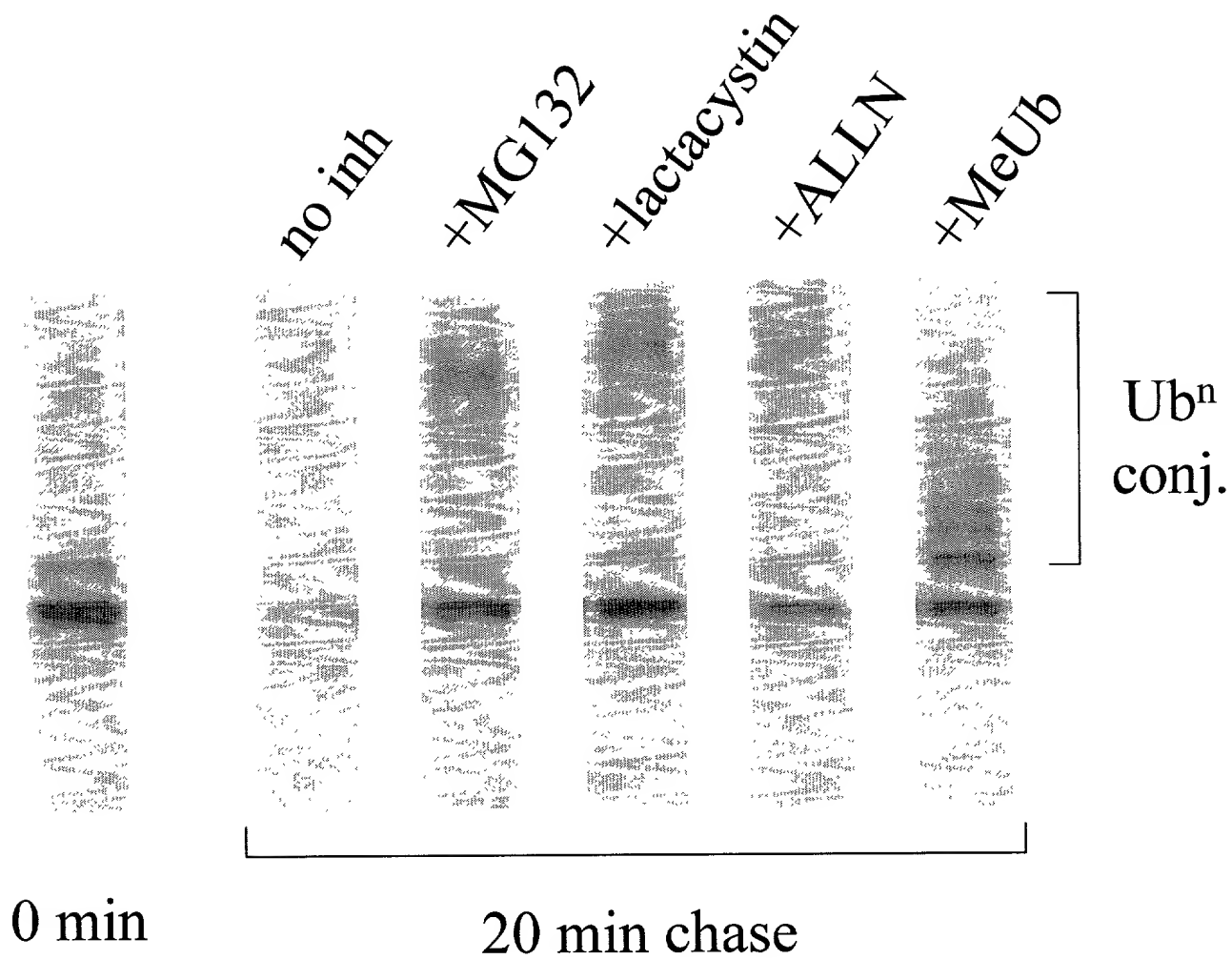


**B.**

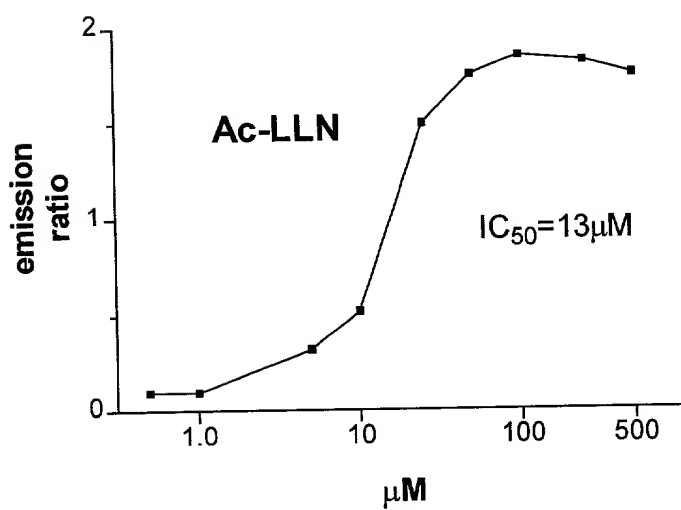
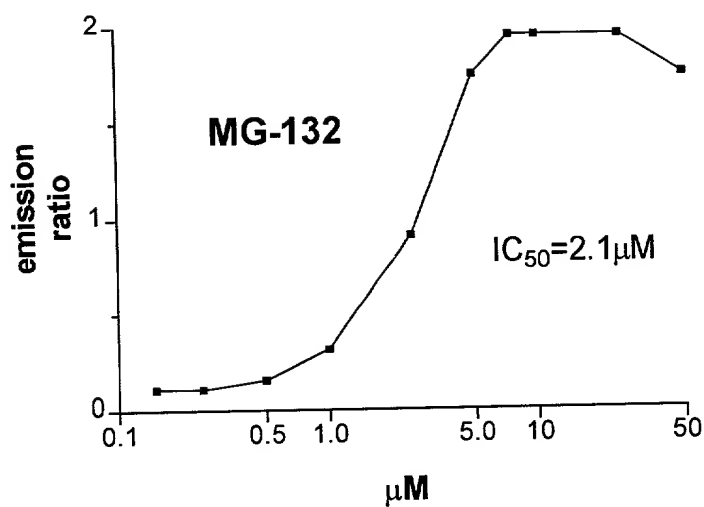
## 3XUb-HRV14-Bla



# Figure 10



# Figure 11



# DECLARATION FOR PATENT APPLICATION

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship is as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF, the specification of which

\_\_\_\_\_ is attached hereto.

X was filed on February 4, 2000 (Attorney Docket No. AURO1330) as U.S. Application Serial No. \_\_\_\_\_ and was amended on \_\_\_\_\_ if applicable (the "Application").

I hereby authorize and request insertion of the application serial number of the Application when officially known.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the Application as defined in Title 37, Code of Federal Regulations ("C.F.R."), § 1.56.

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of the application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the Application as defined in Title 37, C.F.R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of the Application:

<u>(U.S. Patent No.)</u>	<u>(Issue Date)</u>	<u>(Status)</u> (patented, pending, abandoned)
--------------------------	---------------------	---------------------------------------------------

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of first inventor: Jeffrey Stack

Inventor's signature: 

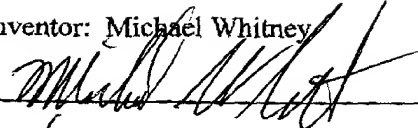
Date: 2/3/00

Residence: San Diego, California

Citizenship: United States

Post Office Address: 6619 Edmonton Avenue, San Diego, CA 92122

Full name of second inventor: Michael Whitney

Inventor's signature: 

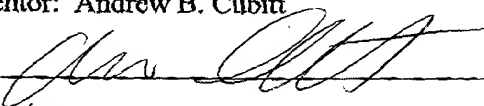
Date: 2/2/00

Residence: San Diego, California

Citizenship: United States

Post Office Address: 5171 Frink Avenue, San Diego, California 92117

Full name of third inventor: Andrew B. Cubitt

Inventor's signature: 

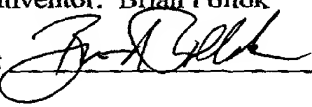
Date: 1/2/00

Residence: San Diego, California

Citizenship: United States

Post Office Address: 11083 Caminito Encanto, San Diego, California 92131

Full name of fourth inventor: Brian Pollok

Inventor's signature: 

Date: 2/3/00

Residence: San Diego, California

Citizenship: United States

Post Office Address: 18283 Sun Maiden Court, San Diego, California 92127

## PATENT

)  
)  
)  
)  
)  
)  
)  
)  
)  
)  
)

[illegible]

**As a below-named assignee of the above-identified application ("Application"):**

I hereby appoint the following attorneys of the assignee to prosecute the Application and to transact all business in the United States Patent and Trademark Office connected therewith:

Registration No. 41,734  
 Registration No. 38,347  
 Registration No. 33,402  
 Registration No. 37,643  
 Registration No. 44,835  
 Registration No. 31,238  
 Registration No. 35,255  
 Registration No. 30,298  
 Registration No. 41,234  
 Registration No. 31,192  
 Registration No. 40,825  
 Registration No. 38,322  
 Registration No. 38,626  
 Registration No. 27,774  
 Registration No. 38,770  
 Registration No. P-45,452  
 Registration No 41,285

GT6166741.1  
103651-990000

RECEIVED TIME FEB. 4. 11:16AM

619 404 6743

In re Application of:

PATENT  
Attorney Docket No.: AURO1330

Application No.:

Filed: Herewith

Page 2

I hereby authorize and request insertion of the application number of the  
Application when officially known.

Direct all telephone calls to:

LISA A. HAILE, PH.D.  
Telephone: (858) 677-1456

Address all correspondence to:

LISA A. HAILE, PH.D.  
GRAY CARY WARE & FREIDENRICH LLP  
4365 Executive Drive, Suite 1600  
San Diego, CA 92121**AURORA BIOSCIENCES CORPORATION**

a corporation of the state of Delaware

By: \_\_\_\_\_

Name: \_\_\_\_\_

John D. Mendlein

Senior Vice President, Intellectual Property

Title: \_\_\_\_\_

Chief Knowledge Officer

Date: \_\_\_\_\_

02-04-2000